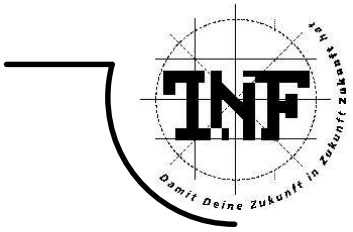
JOHANNES KEPLER
UNIVERSITÄT LINZ
Netzwerk für Forschung, Lehre und Praxis



Informations- und Codierungstheorie

(Informationstheorie)

Vorlesungsunterlagen

Wintersemester 2006/07

Erhard Aichinger
Institut für Algebra
Johannes Kepler Universität Linz

Alle Rechte vorbehalten

Linz, im Jänner 2007

Adresse:
Doz. Dr. Erhard Aichinger
Institut für Algebra
Johannes Kepler Universität Linz
4040 Linz
e-mail: erhard@algebra.uni-linz.ac.at

Druck: Kopierstelle der Uni Linz

Inhaltsverzeichnis

KAPITEL 1

Überblick

1. Einige Begriffe der Informationstheorie

Ein Kommunikationssystem besteht aus folgenden Teilen (s. [**HQ95**, p. 1]).

(1) Nachrichtenquelle
(2) Quellencodierer
(3) Kanalcodierer
(4) Kanal
(5) Rausch- oder Fehlerquelle
(6) Kanaldecodierer
(7) Quellendecodierer
(8) Nachrichtensenke

2. Quellencodierung und Datenkompression

Wir stellen uns vor, wir haben eine Nachrichtenquelle, die jede Sekunde eines der Zeichen A, B, C, D liefert. Dabei sind 60% der Zeichen A, 30% der Zeichen B, 5% der Zeichen C und 5% der Zeichen D. Wir nehmen an, die Nachrichtenquelle produziert die Zeichen unabhängig voneinander; die Auswahl des nächsten Zeichens wird also von den bisher gelieferten Zeichen nicht beeinflusst.

Wir wollen eine von dieser Nachrichtenquelle produzierte Nachricht möglichst effizient als 0/1-Folge speichern.

Als erste Idee ersetzen wir jedes der 4 Zeichen durch eine 0/1 Folge. Wir vergleichen drei verschiedene Vorschläge.

Zeichen	Vorschlag 1	Vorschlag 2	Vorschlag 3
A	00	0	0
B	01	10	10
C	10	110	110
D	11	01	111

Wir überlegen uns, wie lange die 0/1-Folge ist, die wir brauchen, um eine Datei mit n Zeichen aus dieser Nachrichtenquelle zu speichern.

(1) Im Vorschlag 1 brauchen wir $2 \cdot n$ Zeichen für jede Datei mit n Zeichen. Wir brauchen also 2 Bits pro Nachrichtenzeichen.
(2) Der Vorschlag 2 hat folgenden Nachteil: DB und AC werden beide durch 0110 dargestellt. Daher ist ein eindeutiges Decodieren nicht mehr möglich.
(3) Im Vorschlag 3 ist ein eindeutiges Decodieren möglich. Hier werden verschiedene Dateien der Länge n zu möglicherweise unterschiedlich langen Dateien kodiert. Wenn wir eine "typische" Datei wählen, so hat diese Datei etwa $0.6\,n$ mal das Zeichen A, $0.3\,n$ mal das Zeichen B und jeweils $0.05\,n$ mal die Zeichen C und D. Diese Datei wird mit

$$1 \cdot 0.6\,n + 2 \cdot 0.3\,n + 3 \cdot 0.05\,n + 3 \cdot 0.05\,n$$

$= 1.5\,n$ Bits dargestellt. Wir brauchen also 1.5 Bits/Nachrichtenzeichen.

Typische Resultate über die Quellcodierung:

(1) Shannon's Quellcodierungssatz (Noiseless Coding Theorem): Wenn die Zeichen $A_1, A_2, \ldots, A_n$ mit Wahrscheinlichkeiten $p_1, p_2, \ldots, p_n$ auftreten, so braucht jedes Quellcodierungsverfahren, das für beliebig lange Dateien funktioniert, im Mittel zumindest

$$\sum_{i=1}^{n} p_i \cdot \log_2(\frac{1}{p_i})$$

Bits pro Nachrichtenzeichen. Durch geeignete Codierungsverfahren kann man dieser Schranke beliebig nahe kommen. Für das obige Beispiel ist diese Schranke ungefähr 1.395 Bits pro Nachrichtenzeichen.
(2) Huffman-Algorithmus [**Ash90**, p. 42], [**Mac03**, p.99]: Für gegebene Zeichen $A_1, A_2, \ldots, A_n$ mit Wahrscheinlichkeiten $(p_1, p_2, \ldots, p_n)$ produziert der Huffman-Algorithmus die beste zeichenweise Codierung von $A_1, A_2, \ldots, A_n$ als 0/1-Folgen.

3. Kanalcodierung

Wir stellen uns vor, wir haben einen Kanal zur Verfügung, der 0 und 1 überträgt. Dabei kommt das Eingabezeichen 0 mit Wahrscheinlichkeit f (Fehlerwahrscheinlichkeit) als 1 und mit Wahrscheinlichkeit $1 - f$ als 0 an. Ebenso kommt ein gesandter 1er mit Wahrscheinlichkeit f als 0 und mit Wahrscheinlichkeit $1 - f$ als 1 an. Wir suchen nun Codierungsverfahren die so funktioneren:

(1) Eine Folge $x_1x_2x_3x_4\ldots$ von Bits soll übertragen werden.
(2) Der *Kanalcodierer* fügt Kontrollbits dazu und macht aus $x_1x_2x_3x_4$ die Folge $y_1y_2y_3y_4y_5y_6\ldots$.

(3) Diese Folge $y_1y_2y_3y_4y_5y_6\ldots$ wird über den Kanal gesendet. Die Folge $z_1z_2z_3z_4z_5z_6\ldots$ kommt an. Wir nehmen an, dass mit Wahrscheinlichkeit $1-f$ ein z_i mit dem ausgesandten y_i übereinstimmt.
(4) Der *Kanaldecodierer* versucht, die vermutlich ausgesandte Folge $x_1x_2x_3x_4\ldots$ zu rekonstruieren. Er liefert also eine Bitfolge $u_1u_2u_3u_4\ldots$.

Die *Bitfehlerrate* ist die Wahrscheinlichkeit, dass ein Nachrichtenbit x_i nicht mit dem decodierten Bit u_i übereinstimmt.

Die *Übertragungsrate* ist die Anzahl der Nachrichtenbits pro gesendetem Kanalbit. Wir vergleichen zwei Varianten, um eine Bitfolge über einen Kanal zu schicken.

(1) "Keine" Codierung und Decodierung: Wir schicken die Nachricht ohne Kontrollstellen über den Kanal. Wenn f gegeben ist, können wir die Bitfehlerrate b und die Übertragungsrate r bestimmen:

$$b = f \text{ und } r = 1.$$

(2) Wir wiederholen jedes Bit drei Mal und decodieren durch Mehrheitsentscheidung. Es gilt:

$$b = f^3 + 3f^2(1-f) \text{ und } r = \frac{1}{3}.$$

Bei einem Kanal mit Fehlerwahrscheinlichkeit $f = 0.1$ können wir so eine Bitfehlerrate von 0.028 erreichen. Wenn wir durch den gleichen Kanal jedes Bit 1001 mal schicken und dann durch Mehrheitsentscheidung decodieren, erhalten wir die Fehlerwahrscheinlichkeit $b = 8.0276 \cdot 10^{-225}$. Der Preis, den wir dafür bezahlen, ist eine Ver-1001-fachung der Übertragungskosten, die Übertragungsrate ist in diesem Fall ja nur $r = \frac{1}{1001}$.

Wo liegen die theoretischen Grenzen für die Bitfehlerrate b bei gegebener Kanalfehlerrate f und Übertragungsrate r? Shannons Kanalcodierungssatz beantwortet diese Frage (cf. [**Mac03**, p. 15]).

(1) Wenn $r < 1 + f \cdot \log_2(f) + (1-f) \cdot \log_2(1-f)$ ist, und wenn ε eine vorgegebene maximal zulässige Bitfehlerrate mit $\varepsilon > 0$ ist, dann gibt es eine Codierung und Decodierung mit Übertragungsrate r, für die die Bitfehlerrate kleiner als ε ist.

    Der Preis, den wir für diese Sicherheit zahlen müssen, ist, dass wir eventuell lange Blöcke von N Nachrichtenzeichen abwarten müssen, um diese dann als $N \cdot \frac{1}{r}$ Bits zu kodieren. Dann erst senden wir diese $\frac{N}{r}$ Bits über den Kanal.

(2) Wenn $r > 1 + f \cdot \log_2(f) + (1-f) \cdot \log_2(1-f)$, dann ist die bestenfalls erreichbare Bitfehlerrate b jene Lösung der Gleichung

$$r \cdot \big(1 + b \cdot \log_2(b) + (1-b) \cdot \log_2(1-b)\big) = 1 + f \cdot \log_2(f) + (1-f) \cdot \log_2(1-f),$$

für die $b < \frac{1}{2}$ ist. Kleinere Bitfehlerraten als dieses b sind unmöglich. Es gibt Codierungsverfahren (wieder mit Puffern der Nachricht), die diesem b beliebig nahe kommen.

Die Codierungstheorie bemüht sich, praktisch durchführbare Codierungsverfahren zu finden, die bei gegebenem r und f das b möglichst klein werden lassen (siehe [**Wil99**]).

KAPITEL 2

Grundbegriffe der Wahrscheinlichkeitstheorie

1. Problemstellungen der Wahrscheinlichkeitstheorie

Eine Nachrichtenquelle produziert eine Folge von 100 Zeichen auf folgende Art: Für jedes Zeichen wird gewürfelt. Fällt der Würfel auf 1, 2, oder 3, so wird das Zeichen a produziert. Fällt der Würfel auf 4 oder 5, wird b produziert. Fällt der Würfel auf 6, so wird c produziert.

Wir fragen uns nun, wieviele a wir erwarten dürfen. Zunächst würde man erwarten, dass etwa die Hälfte der Zeichen a sind. Dennoch ist es möglich, dass der Würfel 100 mal hintereinander auf 4, 5, oder 6 fällt, und kein einziges Zeichen ein a ist. "Möglich schon, aber sehr unwahrscheinlich", sagt die Wahrscheinlichkeitstheorie dazu.

Produziert man nun anstatt 100 Zeichen 10000 oder 100000 Zeichen, so scheint es plausibel, dass der Anteil an a den erwarteten 50% schließlich ganz nahe kommt. Trotzdem ist es denkbar, dass auch unter 100000 Zeichen kein einziges a vorkommt. In den "Gesetzen der großen Zahlen" erhalten wir mathematische Aussagen darüber.

2. Wahrscheinlichkeit

Wir wollen beschreiben, dass die drei Zeichen a, b, c mit den "Wahrscheinlichkeiten" $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{6}$ auftreten. Dazu definieren wir den Begriff *Wahrscheinlichkeitsraum*. Ein *Wahrscheinlichkeitsraum* besteht aus einer endlichen Menge, und einer Funktion, die jedem Element seine relative Häufigkeit zuordnet.

Definition 2.1 (Endliche Wahrscheinlichkeitsräume). Sei Ω eine endliche nichtleere Menge, und sei $P : \Omega \to [0, 1]$. Das Paar (Ω, P) ist ein *Wahrscheinlichkeitsraum*, falls

$$\sum_{\omega \in \Omega} P(\omega) = 1.$$

Wir können nun jede Teilmenge eines Wahrscheinlichkeitsraumes messen.

DEFINITION 2.2. Sei (Ω, P) ein Wahrscheinlichkeitsraum, und sei A eine Teilmenge von Ω. Dann definieren wir das *Maß von* A, $P(A)$, durch

$$P(A) := \sum_{a \in A} P(a).$$

In der Wahrscheinlichkeitstheorie beschreibt man Vorgänge, die vom Zufall abhängen, etwa vom Wurf einer Münze, von der Augenzahl, die ein Würfel zeigt, oder von der Kiste, in die eine Roulettekugel fällt. So hängt die Auswahl der Zeichen a, b, c im Beispiel der vorigen Sektion davon ab, wie der Würfel gefallen ist. Es ist zufällig, welches der drei Zeichen a, b, c die Nachrichtenquelle als nächstes produziert. Wenn aber der Würfel gefallen ist, so ist die Auswahl bestimmt. Für die mathematische Beschreibung der Auswahl eines Zeichens trennt man das Würfeln vom deterministischen Vorgang, der Auswahl des Zeichens aus der Augenzahl.

Das Würfeln kodiert man durch einen Wahrscheinlichkeitsraum (Ω, P) mit

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

und

$$P(1) = P(2) = \cdots = P(6) = \frac{1}{6}.$$

Die Auswahl des Zeichens kodiert man durch eine Funktion von $\{1, 2, 3, 4, 5, 6\}$ nach $\{a, b, c\}$. Im genannten Beispiel verwenden wir die Funktion X mit $X(1) = X(2) = X(3) = a$, $X(4) = X(5) = b$, $X(6) = c$. Funktionen, die auf der Trägermenge eines Wahrscheinlichkeitsraumes definiert sind, heißen *Zufallsvariablen*.

DEFINITION 2.3. Sei (Ω, P) ein Wahrscheinlichkeitsraum, und sei M eine Menge. Eine M-wertige *Zufallsvariable* X auf (Ω, P) ist eine Funktion $X : \Omega \to M$.

Wir interessieren uns nun dafür, wie häufig das Zeichen b ausgewählt wird. Dazu messen wir, wie groß die Menge $\{\omega \mid X(\omega) = b\}$ ist, nennen das Maß dieser Menge *die Wahrscheinlichkeit für* $X = b$, und kürzen diese mit $P[X = b]$ ab. In unserem Fall erhalten wir

$$P[X = b] = P(\{\omega \mid X(\omega) = b\}) = P(\{4, 5\}) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}.$$

DEFINITION 2.4 (Wahrscheinlichkeit). Sei (Ω, P) ein Wahrscheinlichkeitsraum, M eine Menge, und $X : \Omega \to M$ eine Zufallsvariable. Sei Z eine Teilmenge von M. Dann definieren wir *die Wahrscheinlichkeit, dass* X *in* Z *liegt*, abgekürzt $P[X \in Z]$, durch

$$P[X \in Z] := P(\{\omega \mid X(\omega) \in Z\}).$$

Wenn Z einelementig ist und $Z = \{z\}$, dann schreiben wir auch $P[X = z]$ für $P[X \in Z]$.

Wir überlegen uns nun, wie wir Vorgänge beschreiben, die von mehrmaligem Würfeln abhängen. Dazu betrachten wir folgendes Beispiel: Wie groß ist die Wahrscheinlichkeit, bei zweimaligem Würfeln die Augensumme 6 zu erhalten? Wir gehen davon aus, dass bei zweimaligem Würfeln jede Folge gleich wahrscheinlich ist: $(1,1)$ ist also gleich wahrscheinlich wie $(2,3)$, $(3,2)$ oder $(1,6)$. Daher definieren wir einen Wahrscheinlichkeitsraum (Ω, P) durch

$$\begin{aligned} \Omega &= \{(\omega_1,\omega_2) \mid \omega_1,\omega_2 \in \{1,2,3,4,5,6\}\}, \\ P((\omega_1,\omega_2)) &= \tfrac{1}{36} \text{ für alle } \omega_1,\omega_2 \in \{1,2,3,4,5,6\}. \end{aligned}$$

Die Zufallsvariable X definieren wir durch

$$X((\omega_1,\omega_2)) := \omega_1 + \omega_2.$$

Wir suchen die Wahrscheinlichkeit für $X = 6$. Wir erhalten

$$\begin{aligned} P[X=6] &= P(\{(\omega_1,\omega_2) \in \{1,2,3,4,5,6\}^2 \mid X((\omega_1,\omega_2)) = 6\}) \\ &= P(\{(\omega_1,\omega_2) \in \{1,2,3,4,5,6\}^2 \mid \omega_1+\omega_2 = 6\}) \\ &= P(\{(1,5),(2,4),(3,3),(4,2),(5,1)\}) \\ &= \frac{5}{36}. \end{aligned}$$

Also ist die Augensumme mit Wahrscheinlichkeit $\frac{5}{36}$, also $13.888\ldots\,\%$, gleich 6.

Wir betrachten bei zweimaligem Würfeln folgende Zufallsvariablen:

(1) $X((\omega_1,\omega_2)) := \omega_1$,
(2) $Y((\omega_1,\omega_2)) := \omega_2$,
(3) $Z((\omega_1,\omega_2)) := \omega_1 + \omega_2$.

X liefert also das Ergebnis des ersten Wurfes, Y das Ergebnis des zweiten Wurfes, und Z die Summe der Augenzahlen beider Würfe. Der Ausgang von X und der Ausgang von Y beeinflussen einander nicht; wenn ich auch weiß, dass der erste Wurf 5 ist, so bleiben für den zweiten Wurf doch alle Ausgänge gleich wahrscheinlich. Der Ausgang von X liefert aber Einschränkungen für den Ausgang von Z. Ist der erste Wurf 5, so ist die Summe bestimmt nicht mehr 2 oder 12. Wenn Zufallsvariablen einander nicht beeinflussen, so nennt man sie *unabhängig*.

Definition 2.5. Sei (Ω, P) ein Wahrscheinlichkeitsraum, seien M, N Mengen, und seien $X : \Omega \to M$ und $Y : \Omega \to N$ Zufallsvariablen. X und Y sind *unabhängig*, falls für alle $a \in M$ und $b \in N$ gilt:

$$P[X = a \;\&\; Y = b] = P[X = a] \cdot P[Y = b].$$

Dabei steht $P[X = a \;\&\; Y = b]$ für das Maß $P(\{\omega \in \Omega \mid X(\omega) = a \text{ und } Y(\omega) = b\})$.

Lemma 2.6. *Sei (Ω, P) ein Wahrscheinlichkeitsraum, seien M, N Mengen, und seien $X : \Omega \to M$ und $Y : \Omega \to N$ Zufallsvariablen. Äquivalent sind:*

(1) *X und Y sind unabhängig.*
(2) *Für alle Teilmengen A von M und für alle Teilmengen B von N gilt $P[X \in A \ \& \ Y \in B] = P[X \in A] \cdot P[Y \in B]$.*

Dabei steht $P[X \in A \ \& \ Y \in B]$ für $P(\{\omega \in \Omega \mid X(\omega) \in A$ und $Y(\omega) \in B\})$.

Beweis: (1)⇒(2):

$$\begin{aligned} P[X \in A \ \& \ Y \in B] &= P(\{\omega \in \Omega \mid X(\omega) \in A \text{ und } Y(\omega) \in B\}) \\ &= \sum_{\substack{\omega \in \Omega \\ X(\omega) \in A \text{ und } Y(\omega) \in B}} P(\omega) \\ &= \sum_{a \in A} \sum_{b \in B} \sum_{\substack{\omega \in \Omega \\ X(\omega) = a \text{ und } Y(\omega) = b}} P(\omega) \\ &= \sum_{a \in A} \sum_{b \in B} P(\{\omega \in \Omega \mid X(\omega) = a \text{ und } Y(\omega) = b\}) \\ &= \sum_{a \in A} \sum_{b \in B} P[X = a \ \& \ Y = b]. \end{aligned}$$

Nun verwenden wir die Unabhängigkeit von X und Y und erhalten

$$\begin{aligned} \sum_{a \in A} \sum_{b \in B} P[X = a \ \& \ Y = b] &= \sum_{a \in A} \sum_{b \in B} P[X = a] \cdot P[Y = b] \\ &= (\sum_{a \in A} P[X = a]) \cdot (\sum_{b \in B} P[X = b]) \\ &= P[X \in A] \cdot P[X \in B]. \end{aligned}$$

(2)⇒(1): Wir fixieren $a \in M$ und $b \in N$. Wir erhalten $P[X = a \ \& \ Y = b] = P[X \in \{a\} \ \& \ Y \in \{b\}]$. Wegen (2) ist der letzte Ausdruck gleich $P[X \in \{a\}] \cdot P[Y \in \{b\}] = P[X = a] \cdot P[X = b]$. □

Definition 2.7 (Bedingte Wahrscheinlichkeit). Sei (Ω, P) ein Wahrscheinlichkeitsraum, seien M, N Mengen, und seien $X : \Omega \to M$ und $Y : \Omega \to N$ Zufallsvariablen. Sei A eine Teilmenge von M und B eine Teilmenge von N mit $P[Y \in B] \neq 0$. Dann definieren wir

$$P[X \in A \mid Y \in B] := \frac{P[X \in A \ \& \ Y \in B]}{P[Y \in B]},$$

und nennen die linke Seite *die Wahrscheinlichkeit, dass X in A ist, wenn wir schon wissen, dass Y in B ist.*

Wir schränken also unser Interesse auf jene Ereignisse ein, für die Y in B liegt, und messen, für welchen Anteil von diesen X in A liegt.

Wir überlegen uns dazu folgendes Beispiel: Wie groß ist die Wahrscheinlichkeit, dass bei zweimaligem Würfeln der erste Wurf ein Einser ist, wenn die Augensumme 4 ist?

Zur Lösung wählen wir den Wahrscheinlichkeitsraum

$$\begin{aligned} \Omega &= \{(\omega_1, \omega_2) \mid \omega_1, \omega_2 \in \{1,2,3,4,5,6\}\}, \\ P((\omega_1, \omega_2)) &= \tfrac{1}{36} \text{ für alle } \omega_1, \omega_2 \in \{1,2,3,4,5,6\}. \end{aligned}$$

Wir definieren die Zufallsvariablen X und Y durch

$$\begin{aligned} X(\omega_1, \omega_2) &:= \omega_1 \text{ für alle } \omega_1, \omega_2 \in \{1,2,3,4,5,6\}, \\ Y(\omega_1, \omega_2) &:= \omega_1 + \omega_2 \text{ für alle } \omega_1, \omega_2 \in \{1,2,3,4,5,6\}. \end{aligned}$$

Gesucht ist dann $P[X = 1 \mid Y = 4]$. Wir berechnen

$$\begin{aligned} P[X = 1 \mid Y = 4] &= \frac{P[X = 1 \ \& \ Y = 4]}{P[Y = 4]} \\ &= \frac{P(\{\omega \in \Omega \mid X(\omega) = 1 \text{ und } Y(\omega) = 4\})}{P(\{\omega \in \Omega \mid Y(\omega) = 4\})} \\ &= \frac{P(\{(1,3)\})}{P(\{(1,3),(2,2),(3,1)\})} \\ &= \frac{\frac{1}{36}}{\frac{3}{36}} \\ &= \frac{1}{3}. \end{aligned}$$

Die Wahrscheinlichkeit dafür, dass der erste Wurf ein Einser war, wenn wir schon wissen, dass die Augensumme 4 ist, ist also $\frac{1}{3}$.

Aus Neugier stellen wir auch folgende Frage: Wie groß ist die Wahrscheinlichkeit, dass die Augensumme 4 ist, wenn wir schon wissen, dass der erste Wurf ein Einser war. Wir berechnen dazu

$$\begin{aligned} P[Y = 4 \mid X = 1] &= \frac{P[Y = 4 \ \& \ X = 1]}{P[X = 1]} \\ &= \frac{P(\{(1,3)\})}{P(\{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6)\})} \\ &= \frac{\frac{1}{36}}{\frac{6}{36}} \\ &= \frac{1}{6}. \end{aligned}$$

Die Wahrscheinlichkeit dafür, dass die Augensumme 4 ist, wenn wir schon wissen, dass der erste Wurf ein Einser ist, ist also $\frac{1}{6}$. Wenn wir uns in dieses Experiment erst zu dem Zeitpunkt einschalten, an dem der erste Würfel auf Eins gefallen

ist, müssen wir nur mehr beobachten, ob der zweite Würfel auf 3 fällt, damit die Augensumme 4 ist. Der zweite Würfel fällt mit Wahrscheinlichkeit $\frac{1}{6}$ auf 3.

PROPOSITION 2.8. *Sei (Ω, P) ein Wahrscheinlichkeitsraum, seien M, N Mengen, seien $X : \Omega \to M$ und $Y : \Omega \to N$ Zufallsvariablen , und sei sei $A \subseteq M$, $B \subseteq N$. Wenn X und Y unabhängig sind und $P[Y \in B] \neq 0$, so gilt*

$$P[X \in A \mid Y \in B] = P[X \in A].$$

Selbst wenn wir wissen, dass Y in B liegt, hat das keinen Einfluss auf die Wahrscheinlichkeit dafür, dass X in A liegt.

Beweis:

$$P[X \in A \mid Y \in B] = \frac{P[X \in A \ \& \ Y \in B]}{P[Y \in B]}.$$

Nun verwenden wir, dass X und Y unabhängig sind, und erhalten

$$\begin{aligned} \frac{P[X \in A \ \& \ Y \in B]}{P[Y \in B]} &= \frac{P[X \in A] \cdot P[Y \in B]}{P[Y \in B]} \\ &= P[X \in A]. \end{aligned}$$

□

3. Erwartungswert und Varianz

Wenn X eine Zufallsvariable vom Wahrscheinlichkeitsraum Ω nach $\mathbb{R}$ ist, dann bezeichnen wir mit $E(X)$ den *Durchschnitt* ihrer Ausgänge, und nennen ihn *Erwartungswert*.

DEFINITION 2.9. Sei (Ω, P) ein Wahrscheinlichkeitsraum, und sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable. Dann definieren wir den *Erwartungswert $E(X)$ von X* durch

$$E(X) = \sum_{\omega \in \Omega} X(\omega) \cdot P(\omega).$$

Wie üblich bezeichnen wir $\{X(\omega) \mid \omega \in \Omega\}$ mit $X(\Omega)$.

LEMMA 2.10. *Sei (Ω, P) ein Wahrscheinlichkeitsraum, und sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable. Dann gilt*

$$E(X) = \sum_{s \in X(\Omega)} s \cdot P[X = s]. \tag{2.1}$$

Beweis: Wir definieren $\delta(a,b) = 1$, falls $a = b$ und $\delta(a,b) = 0$, falls $a \neq b$. Wir formen nun die rechte Seite von (2.1) um und erhalten:

$$\begin{aligned}\sum_{s\in X(\Omega)} s \cdot P[X=s] &= \sum_{s\in X(\Omega)} s \cdot P(\{\omega \in \Omega \mid X(\omega) = s\}) \\ &= \sum_{s\in X(\Omega)} s \cdot \sum_{\substack{\omega\in\Omega \\ X(\omega)=s}} P(\omega) \\ &= \sum_{s\in X(\Omega)} s \cdot \sum_{\omega\in\Omega} P(\omega)\cdot\delta(X(\omega),s) \\ &= \sum_{s\in X(\Omega)}\sum_{\omega\in\Omega} s\cdot P(\omega)\cdot\delta(X(\omega),s) \\ &= \sum_{\omega\in\Omega}\sum_{s\in X(\Omega)} s\cdot P(\omega)\cdot\delta(X(\omega),s).\end{aligned}$$

In der inneren Summe im letzten Ausdruck ist nur der Summand mit $s = X(\omega)$ ungleich 0. Daher ist der letzte Ausdruck gleich

$$\sum_{\omega\in\Omega} X(\omega)\cdot P(\omega)\cdot 1,$$

und das ist genau der Erwartungswert $E(X)$. □

Eine wichtige Eigenschaft des Erwartungswertes ist seine Linearität.

Satz 2.11. *Sei (Ω, P) ein Wahrscheinlichkeitsraum, seien X, Y Zufallsvariablen von $\Omega \to \mathbb{R}$, und sei $t \in \mathbb{R}$. Dann gilt:*

(1) $E(X+Y) = E(X) + E(Y)$.
(2) $E(t\cdot X) = t \cdot E(X)$.

Dabei ist $t \cdot X$ die Funktion, die jedes ω auf $t \cdot X(\omega)$ abbildet.

Beweis: Die erste Eigenschaft beweisen wir durch

$$\begin{aligned}E(X+Y) &= \sum_{\omega\in\Omega}(X(\omega)+Y(\omega))\cdot P(\omega) \\ &= \sum_{\omega\in\Omega}(X(\omega)\cdot P(\omega) + Y(\omega)\cdot P(\omega)) \\ &= \sum_{\omega\in\Omega} X(\omega)\cdot P(\omega) + \sum_{\omega\in\Omega} Y(\omega)\cdot P(\omega) \\ &= E(X) + E(Y).\end{aligned}$$

Die zweite Eigenschaft beweisen wir durch

$$\begin{aligned} E(t \cdot X) &= \sum_{\omega \in \Omega} (t \cdot X)(\omega) \cdot P(\omega) \\ &= \sum_{\omega \in \Omega} t \cdot X(\omega) \cdot P(\omega) \\ &= t \cdot \sum_{\omega \in \Omega} X(\omega) \cdot P(\omega) \\ &= t \cdot E(X). \end{aligned}$$

□

Die folgende Eigenschaft unabhängiger Zufallsvariablen ist verblüffend:

LEMMA 2.12. *Sei (Ω, P) ein Wahrscheinlichkeitsraum, und seien $X, Y : \Omega \to \mathbb{R}$ unabhängige Zufallsvariablen. Dann gilt*

$$E(X \cdot Y) = E(X) \cdot E(Y).$$

Beweis: Wir berechnen $E(X) \cdot E(Y)$. Wir verwenden dazu die Darstellung von $E(X)$ aus Lemma 2.10 und erhalten

$$\begin{aligned} E(X) \cdot E(Y) &= \Big(\sum_{r \in X(\Omega)} r \cdot P[X = r] \Big) \cdot \Big(\sum_{s \in Y(\Omega)} s \cdot P[Y = s] \Big) \\ &= \sum_{r \in X(\Omega)} \sum_{s \in Y(\Omega)} r \cdot s \cdot P[X = r] \cdot P[Y = s]). \end{aligned}$$

Nun verwenden wir die Unabhängigkeit der Zufallsvariablen X und Y.

$$\sum_{r \in X(\Omega)} \sum_{s \in Y(\Omega)} r \cdot s \cdot P[X = r] \cdot P[Y = s] = \sum_{r \in X(\Omega)} \sum_{s \in Y(\Omega)} r \cdot s \cdot P[X = r \ \& \ Y = s].$$

Wir verwenden wieder die Funktion δ mit $\delta(a, b) = 1$, falls $a = b$ und $\delta(a, b) = 0$, falls $a \neq b$. Wir fassen für jedes $t \in \mathbb{R}$ alle Summanden zusammen, für die $r \cdot s = t$ ist. So erhalten wir

$$\begin{aligned} &\sum_{r \in X(\Omega)} \sum_{s \in Y(\Omega)} r \cdot s \cdot P[X = r \ \& \ Y = s] \\ &\qquad = \sum_{t \in X(\Omega) \cdot Y(\Omega)} t \cdot \sum_{r \in X(\Omega)} \sum_{s \in Y(\Omega)} P[X = r \ \& \ Y = s] \cdot \delta(r \cdot s, t). \end{aligned}$$

Wir zeigen nun als nächstes:

(2.2) Für alle $t \in \mathbb{R}$: $\displaystyle\sum_{r \in X(\Omega)} \sum_{s \in Y(\Omega)} P[X = r \ \& \ Y = s] \cdot \delta(r \cdot s, t) = P[X \cdot Y = t].$

Um das zu begründen, beweisen wir zunächst:

(2.3) Für alle $t \in \mathbb{R}$ gilt:

$$\bigcup_{r \in X(\Omega)} \bigcup_{s \in Y(\Omega)} \{\omega \in \Omega \mid X(\omega) = r \text{ und } Y(\omega) = s \text{ und } r \cdot s = t\} = \{\omega \in \Omega \mid X(\omega) \cdot Y(\omega) = t\}.$$

Wir fixieren $t \in \mathbb{R}$. Die Inklusion $\subseteq$ ist offensichtlich. Für die Inklusion $\supseteq$ fixieren wir ω in Ω so, dass $X(\omega) \cdot Y(\omega) = t$. Das Element ω kommt dann in jener Menge der Vereinigung auf der linken Seite von (2.3) vor, für die $r := X(\omega)$ und $s := Y(\omega)$ ist. Das beweist (2.3). Da die Mengen, die auf der linken Seite von (2.3) vereinigt werden, paarweise disjunkt sind, gilt

$$\begin{aligned} P\left(\bigcup_{r \in X(\Omega)} \bigcup_{s \in Y(\Omega)} \{\omega \in \Omega \mid X(\omega) = r \text{ und } Y(\omega) = s \text{ und } r \cdot s = t\} \right) \\ = \sum_{r \in X(\Omega)} \sum_{s \in Y(\Omega)} P(\{\omega \in \Omega \mid X(\omega) = r \text{ und } Y(\omega) = s \text{ und } r \cdot s = t\}) \\ = \sum_{r \in X(\Omega)} \sum_{s \in Y(\Omega)} P(\{\omega \in \Omega \mid X(\omega) = r \text{ und } Y(\omega) = s\}) \cdot \delta(r \cdot s, t). \end{aligned}$$

Das Maß der Menge auf der linken Seite von (2.3) steht also auf der linken Seite von (2.2), und das Maß der Menge auf der rechten Seite von (2.3) steht auf der rechten Seite von (2.2). Das beweist (2.2). Wir verwenden nun (2.2), um weiterzurechnen, und erhalten

$$\begin{aligned} \sum_{t \in X(\Omega) \cdot Y(\Omega)} t \cdot \sum_{r \in X(\Omega)} \sum_{s \in Y(\Omega)} P[X = r \;\&\; Y = s] \cdot \delta(r \cdot s, t) &= \sum_{t \in X(\Omega) \cdot Y(\Omega)} t \cdot P[X \cdot Y = t] \\ &= E(X \cdot Y). \end{aligned}$$

□

Ein Maß für die Schwankung einer Zufallsvariablen um ihren Erwartungswert ist die *Varianz*. Für einen Wahrscheinlichkeitsraum (Ω, P) und eine reelle Zahl a bezeichnen wir mit $\overline{a}$ die Funktion, die durch

$$\begin{array}{rccc} \overline{a} & : & \Omega & \longrightarrow & \mathbb{R} \\ & & \omega & \longmapsto & a \end{array}$$

gegeben ist; $\overline{a}$ ist also die konstante Funktion mit Funktionswert a.

Definition 2.13 (Varianz). Sei (Ω, P) ein Wahrscheinlichkeitsraum, und sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable. Dann ist die *Varianz* von X, abgekürzt $V(X)$, definiert durch

$$V(X) := E\left(\left(X - \overline{E(X)} \right)^2 \right).$$

Die Varianz ist also der Erwartungswert der Zufallsvariablen Y, die durch $Y(\omega) = (X(\omega) - E(X))^2$ definiert ist.

SATZ 2.14. *Sei (Ω, P) ein Wahrscheinlichkeitsraum, sei $n \in \mathbb{N}$, und seien $X_1, X_2, \ldots, X_n : \Omega \to \mathbb{R}$ Zufallsvariablen, die paarweise voneinander unabhängig sind. Dann gilt*

$$V(X_1 + X_2 + \cdots + X_n) = V(X_1) + V(X_2) + \cdots + V(X_n).$$

Beweis:

$$\begin{aligned}
V\left(X_1 + X_2 + \cdots + X_n\right) &= E\left(\left(\left(\sum_{i=1}^{n} X_i\right) - \overline{E\left(\sum_{i=1}^{n} X_i\right)}\right)^2\right) \\
&= E\left(\left(\left(\sum_{i=1}^{n} X_i\right) - \overline{\sum_{i=1}^{n} E\left(X_i\right)}\right)^2\right) \\
&= E\left(\left(\left(\sum_{i=1}^{n} X_i\right) - \sum_{i=1}^{n} \overline{E\left(X_i\right)}\right)^2\right) \\
&= E\left(\left(\sum_{i=1}^{n} \left(X_i - \overline{E\left(X_i\right)}\right)\right)^2\right) \\
&= E\left(\sum_{i=1}^{n}\sum_{j=1}^{n} \left(X_i - \overline{E\left(X_i\right)}\right) \cdot \left(X_j - \overline{E\left(X_j\right)}\right)\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} E\left(\left(X_i - \overline{E\left(X_i\right)}\right) \cdot \left(X_j - \overline{E\left(X_j\right)}\right)\right).
\end{aligned}$$

Wir trennen nun die Summanden, für die $i = j$ ist, von denen, für die $i \neq j$. Wir erhalten

$$\begin{aligned}
&\sum_{i=1}^{n}\sum_{j=1}^{n} E\left(\left(X_i - \overline{E\left(X_i\right)}\right) \cdot \left(X_j - \overline{E\left(X_j\right)}\right)\right) \qquad (2.4) \\
&= \sum_{i=1}^{n} E\left(\left(X_i - \overline{E\left(X_i\right)}\right)^2\right) + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} E\left(\left(X_i - \overline{E\left(X_i\right)}\right) \cdot \left(X_j - \overline{E\left(X_j\right)}\right)\right).
\end{aligned}$$

Wir fixieren nun $i, j \in \{1, 2, \ldots, n\}$ mit $i \neq j$ und betrachten den Ausdruck

$$E\left(\left(X_i - \overline{E\left(X_i\right)}\right) \cdot \left(X_j - \overline{E\left(X_j\right)}\right)\right). \qquad (2.5)$$

Wir rechnen

$$E\left(\left(X_i - \overline{E\left(X_i\right)}\right) \cdot \left(X_j - \overline{E\left(X_j\right)}\right)\right) = E\left(X_i \cdot X_j - \overline{E\left(X_i\right)} \cdot X_j - X_i \cdot \overline{E\left(X_j\right)} + \overline{E\left(X_i\right)} \cdot \overline{E\left(X_j\right)}\right).$$

Wir verwenden die Linearität des Erwartungswertes; außerdem verwenden wir, dass $\overline{E\left(X_i\right)}$ und $\overline{E\left(X_j\right)}$ konstante Funktionen sind und somit $E\left(\overline{E\left(X_i\right)} \cdot X_j\right) = E\left(E\left(X_i\right) \cdot X_j\right) = E\left(X_i\right) \cdot E\left(X_j\right)$ gilt. Außerdem verwenden wir, dass der Erwartungswert einer konstanten Funktion gleich ihrem (einzigen) Funktionswert ist.

$$\begin{aligned} &E\left(X_i \cdot X_j - \overline{E\left(X_i\right)} \cdot X_j - X_i \cdot \overline{E\left(X_j\right)} + \overline{E\left(X_i\right)} \cdot \overline{E\left(X_j\right)}\right) \\ &\quad = E\left(X_i \cdot X_j\right) - E\left(X_i\right) \cdot E\left(X_j\right) - E\left(X_j\right) \cdot E\left(X_i\right) + E\left(X_i\right) \cdot E\left(X_j\right) \\ &\quad = E\left(X_i \cdot X_j\right) - E\left(X_i\right) \cdot E\left(X_j\right). \end{aligned}$$

Da X_i und X_j unabhängig sind, liefert Lemma 2.12, dass der letzte Ausdruck 0 ist. Somit ist (2.5) gleich 0. Wenn wir nun in (2.4) weiterrechnen, erhalten wir

$$\begin{aligned} &\sum_{i=1}^{n} E\left(\left(X_i - \overline{E\left(X_i\right)}\right)^2\right) + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} E\left(\left(X_i - \overline{E\left(X_i\right)}\right) \cdot \left(X_j - \overline{E\left(X_j\right)}\right)\right) \\ &\quad = \sum_{i=1}^{n} E\left(\left(X_i - \overline{E\left(X_i\right)}\right)^2\right) = \sum_{i=1}^{n} V\left(X_i\right). \end{aligned}$$

□

4. Das schwache Gesetz der großen Zahlen

Zur Beschreibung von Zufallsexperimenten, bei denen mehrmals hintereinander gewürfelt wird, eignen sich *Produkträume*. Wir kürzen Vektoren mit fettgedruckten Buchstaben ab, also zum Beispiel $(\omega_1, \omega_2, \ldots, \omega_n)$ mit $\boldsymbol{\omega}$.

Definition 2.15 (Produktraum). Sei (Ω, P) ein Wahrscheinlichkeitsraum. Der *n-fache Produktraum* von (Ω, P), abgekürzt $(\Omega, P)^n$, ist das Paar $\left(\Omega^n, P^{(n)}\right)$, das durch

$$\begin{aligned} \Omega^n &= \{\boldsymbol{\omega} \mid \omega_1, \omega_2, \ldots, \omega_n \in \Omega\}, \\ P^{(n)}\left(\boldsymbol{\omega}\right) &= P\left(\omega_1\right) \cdot P\left(\omega_2\right) \cdots P\left(\omega_n\right) \end{aligned}$$

gegeben ist.

Tatsächlich ist auch im Produktraum die Summe der Wahrscheinlichkeiten wieder 1.

PROPOSITION 2.16. *Sei (Ω, P) ein Wahrscheinlichkeitsraum, und sei $n \in \mathbb{N}$. Dann gilt*

$$\sum_{\boldsymbol{\omega} \in \Omega^n} P(\omega_1) \cdot P(\omega_2) \cdots P(\omega_n) = 1. \tag{2.6}$$

Beweis: Wir beweisen die Gleichheit (2.6) mit Induktion nach n. Für $n = 1$ folgt (2.6) unmittelbar aus der Tatsache, dass sich in Ω die Maße der Elemente zu 1 addieren. Sei nun $n \in \mathbb{N}$, $n \geq 2$. Es gilt

$$\begin{aligned}
&\sum_{\boldsymbol{\omega} \in \Omega^n} P(\omega_1) \cdot P(\omega_2) \cdots P(\omega_n) \\
&\qquad = \sum_{\omega_n \in \Omega} \sum_{\boldsymbol{\omega} \in \Omega^{n-1}} P(\omega_1) \cdot P(\omega_2) \cdots P(\omega_{n-1}) \cdot P(\omega_n) \\
&\qquad\qquad = \sum_{\omega_n \in \Omega} P(\omega_n) \cdot \sum_{\boldsymbol{\omega} \in \Omega^{n-1}} P(\omega_1) \cdot P(\omega_2) \cdots P(\omega_{n-1}).
\end{aligned}$$

Wir verwenden die Induktionsvoraussetzung und erhalten

$$\sum_{\omega_n \in \Omega} P(\omega_n) \cdot \sum_{\boldsymbol{\omega} \in \Omega^{n-1}} P(\omega_1) \cdot P(\omega_2) \cdots P(\omega_{n-1}) = \sum_{\omega_n \in \Omega} P(\omega_n) \cdot 1.$$

Da Ω ein Wahrscheinlichkeitksraum ist, ist die Summe der Maße der Elemente in Ω gleich 1. □

Wir überlegen uns, wie groß die Wahrscheinlichkeit ist, bei dreimaligem Würfeln zumindest einen Sechser zu würfeln. Wir setzen

$$\begin{aligned}
\Omega &= \{1, 2, 3, 4, 5, 6\} \\
\Omega^3 &= \{(1,1,1), (1,1,2), \ldots, (6,6,6)\}, \\
P^{(3)}((\omega_1, \omega_2, \omega_3)) &= \left(\tfrac{1}{6}\right)^3 = \tfrac{1}{216}.
\end{aligned}$$

Die Zufallsvariable X, die uns interessiert, definieren wir so:

$$X((\omega_1, \omega_2, \omega_3)) := \text{ die Anzahl der Sechser in } (\omega_1, \omega_2, \omega_3).$$

Dann suchen wir also

$$P^{(3)}\left(\{\boldsymbol{\omega} \in \Omega^3 \mid X(\omega) \geq 1\}\right).$$

Da die Summe der Maße aller Elemente von Ω^3 gleich 1 ist, kann man diese Wahrscheinlichkeit auch durch den Ausdruck

$$1 - P^{(3)}\left(\{\boldsymbol{\omega} \in \Omega^3 \mid X(\omega) = 0\}\right)$$

berechnen. Klarerweise ist $X(\omega)$ genau dann 0, wenn ω in $\{1,2,3,4,5\}^3$ liegt. Es gilt also

$$\begin{aligned} 1 - P^{(3)}\left(\{\boldsymbol{\omega} \in \Omega^3 \mid X(\omega) = 0\}\right) &= 1 - P^{(3)}\left(\{1,2,3,4,5\}^3\right) \\ &= 1 - 5^3 \cdot \frac{1}{216} \\ &\approx 0.4213 \end{aligned}$$

Die Wahrscheinlichkeit dafür, dass mindestens ein Sechser gewürfelt wird, ist also ungefähr 42%.

Wir beschäftigen uns jetzt mit der Frage, wie es bei längerem Würfeln mit der Anzahl der Sechser aussieht. Man sollte ja erwarten, dass etwa ein Sechstel der Würfe Sechser werden. Ein mathematische Aussage darüber finden wir im "schwachen Gesetz der großen Zahlen". Zuvor brauchen wir aber noch ein harmloses Lemma.

LEMMA 2.17 (Čebyšëv, Tschebyschow, Tschebyscheff). *Sei (Ω, P) ein Wahrscheinlichkeitsraum, sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable, und sei $\varepsilon \in \mathbb{R}$ mit $\varepsilon > 0$. Dann gilt*

$$P(\{\omega \mid |X(\omega)| \geq \varepsilon\}) \leq \frac{E(X^2)}{\varepsilon^2}.$$

Wir berechnen $E(X^2)$.

$$\begin{aligned} E\left(X^2\right) &= \sum_{\omega \in \Omega} P(\omega) \left(X(\omega)\right)^2 \\ &= \sum_{\substack{\omega \in \Omega \\ |X(\omega)| \geq \varepsilon}} P(\omega) \left(X(\omega)\right)^2 + \sum_{\substack{\omega \in \Omega \\ |X(\omega)| < \varepsilon}} P(\omega) \left(X(\omega)\right)^2. \end{aligned}$$

Nun wissen wir, dass die Summanden in der ersten Summe mindestens die Größe $P(\omega) \cdot \varepsilon^2$ haben, und dass die Summanden der zweiten Summe positiv sind. Insgesamt erhalten wir

$$\begin{aligned} \sum_{\substack{\omega \in \Omega \\ |X(\omega)| \geq \varepsilon}} P(\omega) \left(X(\omega)\right)^2 + \sum_{\substack{\omega \in \Omega \\ |X(\omega)| < \varepsilon}} P(\omega) \left(X(\omega)\right)^2 &\geq \sum_{\substack{\omega \in \Omega \\ |X(\omega)| \geq \varepsilon}} P(\omega)\, \varepsilon^2 + 0 \\ &= \varepsilon^2 \cdot \sum_{\substack{\omega \in \Omega \\ |X(\omega)| \geq \varepsilon}} P(\omega) = \varepsilon^2 \cdot P(\{\omega \mid |X(\omega)| \geq \varepsilon\}). \end{aligned}$$

Daher gilt

$$\frac{E(X^2)}{\varepsilon^2} \geq P(\{\omega \mid |X(\omega)| \geq \varepsilon\}).$$

□

SATZ 2.18 (Schwaches Gesetz der großen Zahlen). *Sei* (Ω, P) *ein endlicher Wahrscheinlichkeitsraum, sei* $X : \Omega \to \mathbb{R}$ *eine Zufallsvariable, und sei* $\varepsilon > 0$. *Sei* E_n *gegeben durch*

$$E_n := P^{(n)}\left(\{\boldsymbol{\omega} \mid \left|\left(\frac{1}{n}\sum_{i=1}^{n} X(\omega_i)\right) - E(X)\right| \geq \varepsilon\}\right).$$

Dann gilt

$$E_n \leq \frac{V(X)}{\varepsilon^2 \cdot n}.$$

Die Wahrscheinlichkeit, dass bei n Würfen zwischen 16% und 17% Sechser fallen, konvergiert also gegen 1.

Beweis: Sei $n \in \mathbb{N}$. Wir definieren auf dem Produktraum $(\Omega, P)^n$ die Zufallsvariablen

$$\begin{array}{rcl} X_1(\boldsymbol{\omega}) & := & X(\omega_1) \\ X_2(\boldsymbol{\omega}) & := & X(\omega_2) \\ & \vdots & \\ X_i(\boldsymbol{\omega}) & := & X(\omega_i) \text{ für } i \in \{1, 2, \ldots, n\} \\ & \vdots & \\ X_n(\boldsymbol{\omega}) & := & X(\omega_n) \end{array}$$

und

$$Y(\boldsymbol{\omega}) := \left(\frac{1}{n}\sum_{i=1}^{n} X(\omega_i)\right) - E(X).$$

Wir beobachten zunächst, dass für alle i die Gleichheiten $E(X_i) = E(X)$ und $V(X_i) = V(X)$ gelten. Außerdem sind X_i und X_j für $i \neq j$ unabhängig. Wir

berechnen nun den Erwartungswert von Y^2. Es gilt

$$\begin{aligned}
E\left(Y^2\right) &= \sum_{\boldsymbol{\omega}\in\Omega^n} P^{(n)}\left(\boldsymbol{\omega}\right)\cdot\left(Y\left(\boldsymbol{\omega}\right)\right)^2 \\
&= \sum_{\boldsymbol{\omega}\in\Omega^n} P^{(n)}\left(\boldsymbol{\omega}\right)\cdot\left(\frac{1}{n}\left(\sum_{i=1}^{n} X_i\left(\omega\right)\right)-\frac{1}{n}\left(\sum_{i=1}^{n} E\left(X\right)\right)\right)^2 \\
&= \sum_{\boldsymbol{\omega}\in\Omega^n} P^{(n)}\left(\boldsymbol{\omega}\right)\cdot\frac{1}{n^2}\cdot\left(\left(\sum_{i=1}^{n} X_i\left(\omega\right)\right)-\left(\sum_{i=1}^{n} E\left(X_i\right)\right)\right)^2 \\
&= \frac{1}{n^2}\cdot\sum_{\boldsymbol{\omega}\in\Omega^n} P^{(n)}\left(\boldsymbol{\omega}\right)\cdot\left(\left(\sum_{i=1}^{n} X_i\left(\omega\right)\right)-E\left(\sum_{i=1}^{n} X_i\right)\right)^2 \\
&= \frac{1}{n^2}\cdot E\left(\left(\sum_{i=1}^{n} X_i-\overline{E\left(\sum_{i=1}^{n} X_i\right)}\right)^2\right) \\
&= \frac{1}{n^2}\cdot V\left(\sum_{i=1}^{n} X_i\right).
\end{aligned}$$

Für $i\neq j$ sind die Zufallsvariablen X_i und X_j unabhängig. Satz 2.14 ergibt also

$$\begin{aligned}
\frac{1}{n^2}\cdot V\left(\sum_{i=1}^{n} X_i\right) &= \frac{1}{n^2}\sum_{i=1}^{n} V\left(X_i\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} V\left(X\right) \\
&= \frac{1}{n^2}\cdot n\cdot V\left(X\right) \\
&= \frac{1}{n}V\left(X\right).
\end{aligned}$$

Aus dem Lemma von Čebyšëv, Lemma 2.17, wissen wir, dass

$$P^{(n)}\left(\left\{\boldsymbol{\omega}\mid\left|Y\left(\boldsymbol{\omega}\right)\right|\geq\varepsilon\right\}\right)\leq\frac{E\left(Y^2\right)}{\varepsilon^2}.$$

Es gilt also

$$P^{(n)}\left(\left\{\boldsymbol{\omega}\mid\left|\left(\frac{1}{n}\sum_{i=1}^{n} X\left(\omega_i\right)\right)-E\left(X\right)\right|\geq\varepsilon\right\}\right)\leq\frac{\frac{1}{n}V\left(X\right)}{\varepsilon^2},$$

und somit

$$E_n\leq\frac{V\left(X\right)}{\varepsilon^2\cdot n}.$$

5. Konstruktion von Zufallsvariablen

Wir konstruieren nun aus der Zufallsvariable X, die die Auswahl eines Zeichens auf Basis eines Wurfes mit einem Würfel beschreibt, eine Zufallsvariable X^n, die die Auswahl von n Zeichen durch n-maliges Würfeln beschreibt.

DEFINITION 2.19. Sei (Ω, P) ein Wahrscheinlichkeitsraum, sei A eine Menge, sei $n \in \mathbb{N}$ und sei $X : \Omega \to A$ eine Zufallsvariable. Mit $X^{[n]}$ bezeichnen wir die Zufallsvariable, die durch

$$\begin{array}{rccl} X^{[n]} & : & \Omega^n & \longrightarrow & A^n \\ & & \boldsymbol{\omega} & \longmapsto & (X(\omega_1), X(\omega_2), \ldots, X(\omega_n)) \end{array}$$

definiert ist.

Wenn keine Verwechslung möglich ist, schreiben wir auch X^N für $X^{[N]}$.

Die folgende Konstruktion brauchen wir, um die Ergebnisse zweier Zufallsvariablen gemeinsam zu untersuchen.

DEFINITION 2.20. Sei (Ω, P) ein Wahrscheinlichkeitsraum, seien M, N Mengen, und seien $X : \Omega \to M$, $Y : \Omega \to N$ Zufallsvariablen. Dann definieren wir die Zufallsvariable $X \otimes Y$ durch

$$\begin{array}{rccl} X \otimes Y & : & \Omega & \longrightarrow & M \times N \\ & & \omega & \longmapsto & (X(\omega), Y(\omega)). \end{array}$$

KAPITEL 3

Quellcodierung

1. Motivation

Wir versuchen in diesem Kapitel eine Größe zu definieren, die misst, wieviele Bits wir dazu brauchen, den Ausgang einer Zufallsvariablen zu übertragen.

DEFINITION 3.1. Sei A eine Menge. Wir bezeichnen die Menge $\bigcup_{n\in\mathbb{N}} A^n$ mit A^+.

A^+ ist also die Menge aller endlichen Folgen von Elementen aus A. Wir nennen A^+ auch die Menge aller *Wörter* über dem *Alphabet* A.

DEFINITION 3.2. Sei A eine Menge, und sei $W \in A^+$. Die *Länge von* W ist jenes n, für das $W \in A^n$ gilt. Wir kürzen die Länge von W mit $\mathsf{L}(W)$ oder $|W|$ ab.

DEFINITION 3.3. Sei (Ω, P) ein Wahrscheinlichkeitsraum, und sei $X : \Omega \to A$ eine Zufallsvariable. Eine injektive Funktion $C : A^+ \to \{0,1\}^+$ heißt *gültige Codierung der Ausgänge von* X. Wir definieren die *Kompressionsrate von* C durch

$$R\ (C) := \liminf_{n\to\infty} \frac{1}{n} \sum_{\mathbf{a}\in A^n} \underbrace{P[X = a_1] \cdot P[X = a_2] \cdot \cdots \cdot P[X = a_n]}_{=P[X^{[n]}=\mathbf{a}]} \cdot \mathsf{L}(C\,(\mathbf{a})).$$

Dann definieren wir den Informationsgehalt U_1 von X durch

$$U_1\,(X) := \inf\{R\,(C) \mid C \text{ ist gültige Codierung der Ausgänge von } X\}.$$

Die Einheit von $U_1\,(X)$ ist Bits/Ausgang von X.

Wir können $R\,(C)$ auch durch $\liminf_{n\to\infty} \frac{1}{n} \cdot E\left(\mathsf{L} \circ C \circ X^{[n]}\right)$ abkürzen.

Eine sinnvolle Definition des Informationsgehaltes eines Ausganges von X wäre auch die folgende.

DEFINITION 3.4. Sei (Ω, P) ein Wahrscheinlichkeitsraum, und sei $X : \Omega \to A$ eine Zufallsvariable, und sei $n \in \mathbb{N}$. Eine injektive Funktion $C : A^n \to \{0,1\}^+$ heißt *gültige Codierung von* n *Ausgängen von* X. Wir definieren die *Kompressionsrate von* C durch

$$R\,(C) := \frac{1}{n} \sum_{\mathbf{a}\in A^n} P[X^{[n]} = \mathbf{a}] \cdot \mathsf{L}(C\,(\mathbf{a})).$$

Dann definieren wir den Informationsgehalt U_2 von X durch

$$U_2(X) := \liminf_{n\to\infty} \inf\{R(C) \mid C \text{ ist gültige Codierung von } n \text{ Ausgängen von } X\}.$$

Die Einheit von $U_2(X)$ ist wieder Bits/Ausgang von X.

Wir werden sehen, dass sich sowohl $U_1(X)$ als auch $U_2(X)$ leicht berechnen lassen.

2. Entropie

Wir definieren zunächst eine Funktion $h : [0,1] \to \mathbb{R}$ durch

$$\begin{array}{rcccl} h & : & [0,1] & \longrightarrow & \mathbb{R} \\ & & p & \longmapsto & \begin{cases} 0, & \text{wenn } p = 0, \\ -p \cdot \log_2(p) = p \cdot \log_2\left(\frac{1}{p}\right) & \text{sonst.} \end{cases} \end{array}$$

LEMMA 3.5. *Die Funktion h ist auf $[0,1]$ stetig und konkav.*

DEFINITION 3.6 (Entropie). Sei $M \in \mathbb{N}$, und seien $p_1, p_2, \ldots, p_M \in [0,1]$ so, dass $\sum_{i=1}^{M} p_i = 1$. Wir definieren $H(p_1, p_2, \ldots, p_M)$ durch

$$H(p_1, p_2, \ldots, p_M) := \sum_{i=1}^{M} h(p_i) = \sum_{\substack{i=1 \\ p_i \neq 0}}^{M} p_i \log_2(\frac{1}{p_i}).$$

DEFINITION 3.7 (Entropie einer Zufallsvariablen). Sei (Ω, P) ein Wahrscheinlichkeitsraum, sei A eine Menge, und sei $X : \Omega \to A$ eine Zufallsvariable. Dann definieren wir $H(X)$ durch

$$H(X) := \sum_{a \in X(\Omega)} h(P[X=a]) = \sum_{\substack{a \in A \\ P[X=a]>0}} -P[X=a] \cdot \log_2(P[X=a]).$$

LEMMA 3.8. *Für alle $x \in \mathbb{R}$ mit $x > 0$ gilt $\log(x) \leq x - 1$. Gleichheit gilt nur, wenn $x = 1$.*

LEMMA 3.9 (Die Ungleichung von Gibbs). *Seien $M \in \mathbb{N}$, $p_1, p_2, \ldots, p_M \in [0,1]$, und $q_1, q_2, \ldots, q_M \in [0,1]$ so, dass $\sum_{i=1}^{M} p_i = \sum_{i=1}^{M} q_i = 1$. Wir nehmen an, dass für alle $i \in \{1, 2, \ldots, M\}$ mit $q_i = 0$ auch $p_i = 0$ gilt. Dann gilt:*

(1)
$$\sum_{\substack{i=1 \\ p_i \neq 0}}^{M} -p_i \cdot \log(p_i) \leq \sum_{\substack{i=1 \\ q_i \neq 0}}^{M} -p_i \cdot \log(q_i).$$

(2) *Wenn*

$$\sum_{\substack{i=1\\p_i\neq 0}}^{M} -p_i \cdot \log(p_i) = \sum_{\substack{i=1\\q_i\neq 0}}^{M} -p_i \cdot \log(q_i),$$

so gilt für alle $i \in \{1, 2, \ldots, M\}$*:* $p_i = q_i$.

Beweis: (1): Für alle $i \in \{1, 2, \ldots, M\}$ mit $p_i \neq 0$ gilt nach Lemma 3.8

$$\log\left(\frac{q_i}{p_i}\right) \leq \frac{q_i}{p_i} - 1.$$

Daher gilt

$$\begin{aligned}\sum_{\substack{i=1\\p_i\neq 0}}^{M} p_i \log\left(\frac{q_i}{p_i}\right) &\leq \sum_{\substack{i=1\\p_i\neq 0}}^{M} (q_i - p_i)\\ &= \sum_{\substack{i=1\\p_i\neq 0}}^{M} q_i - \sum_{\substack{i=1\\p_i\neq 0}}^{M} p_i\\ &\leq 1 - 1 = 0.\end{aligned}$$

Daher gilt

$$\sum_{\substack{i=1\\p_i\neq 0}}^{M} -p_i \log(p_i) \leq \sum_{\substack{i=1\\p_i\neq 0}}^{M} -p_i \log(q_i).$$

Die rechte Seite ist gleich

$$\sum_{\substack{i=1\\q_i\neq 0}}^{M} -p_i \log(q_i).$$

(2): Wir nehmen an, dass Gleichheit gilt. Dann muss in beiden $\leq$ im Beweis von (1) Gleichheit gelten. Aus dem ersten $\leq$ erhalten wir:

$$\text{Für alle } i \in \{1, 2, \ldots, M\} \text{ mit } p_i \neq 0 \text{ gilt } p_i = q_i.$$

Aus dem zweiten $\leq$ erhalten wir:

$$\sum_{\substack{i=1\\p_i\neq 0}}^{M} q_i = 1. \tag{3.1}$$

Da sich alle q_i zu 1 summieren, müssen also die q_i, die nicht in der Summe in (3.1) vorkommen, 0 sein. Das bedeutet:

$$\text{Für alle } i \in \{1, 2, \ldots, M\} \text{ mit } p_i = 0 \text{ gilt } q_i = 0.$$

Also ist auch für diese i stets $p_i = q_i$. □

KOROLLAR 3.10. *Sei $M \in \mathbb{N}$ und seien $p_1, p_2, \ldots, p_M$ so, dass $\sum_{i=1}^{M} p_i = 1$. Dann gilt*

$$H(p_1, p_2, \ldots, p_M) \leq \log_2(M).$$

KOROLLAR 3.11. *Sei (Ω, P) ein Wahrscheinlichkeitsraum, und seien X und Y auf Ω definierte Zufallsvariablen. Dann gilt $H(X \otimes Y) \leq H(X) + H(Y)$. Die Gleichheit $H(X \otimes Y) = H(X) + H(Y)$ gilt genau dann, wenn X und Y unabhängig sind.*

Beweis: Sei $\{x_1, \ldots, x_M\}$ der Bildbereich von X und $\{y_1, \ldots, y_N\}$ der Bildbereich von Y. Wir schreiben $p(x_i)$ für $P[X = x_i]$, $p(y_j)$ für $P[Y = y_j]$ und $p(x_i, y_j)$ für $P[X = x_i \,\&\, Y = y_j]$. Dann gilt:

$$H(X \otimes Y) = \sum_{\substack{(i,j) \in \{1,\ldots,M\} \times \{1,\ldots,N\} \\ p(x_i, y_j) \neq 0}} -p(x_i, y_j) \cdot \log_2(p(x_i, y_j)).$$

Ebenso gilt

$$\begin{aligned} H(X) + H(Y) &= \sum_{\substack{i=1 \\ p(x_i) \neq 0}}^{M} -p(x_i) \log_2(p(x_i)) \; + \sum_{\substack{j=1 \\ p(y_j) \neq 0}}^{N} -p(y_j) \log_2(p(y_j)) \\ &= \sum_{\substack{i=1 \\ p(x_i) \neq 0}}^{M} \left(-\log_2(p(x_i)) \cdot \sum_{\substack{j=1 \\ p(y_j) \neq 0}}^{N} p(x_i, y_j) \right) \\ &\quad + \sum_{\substack{j=1 \\ p(y_j) \neq 0}}^{N} \left(-\log_2(p(y_j)) \cdot \sum_{\substack{i=1 \\ p(x_i) \neq 0}}^{M} p(x_i, y_j) \right) \\ &= \sum_{\substack{(i,j) \in \{1,\ldots,M\} \times \{1,\ldots,N\} \\ p(x_i) \neq 0 \text{ und } p(y_j) \neq 0}} -p(x_i, y_j) \cdot \log_2(p(x_i) \cdot p(y_j)). \end{aligned}$$

Nach der Ungleichung von Gibbs (Lemma 3.9) gilt nun

$$\sum_{\substack{(i,j)\in\{1,\dots,M\}\times\{1,\dots,N\}\\ p(x_i,y_j)\neq 0}} -p\left(x_i,y_j\right)\cdot\log_2\left(p\left(x_i,y_j\right)\right) \tag{3.2}$$

$$\leq \sum_{\substack{(i,j)\in\{1,\dots,M\}\times\{1,\dots,N\}\\ p(x_i)\neq 0 \text{ und } p(y_j)\neq 0}} -p\left(x_i,y_j\right)\cdot\log_2\left(p\left(x_i\right)\cdot p\left(y_j\right)\right).$$

Gleichheit gilt nach Lemma 3.9 (2) genau dann, wenn für alle $i,j\in\{1,2,\dots,M\}\times\{1,\dots,N\}$ gilt, dass $p\left(x_i,y_j\right)=p\left(x_i\right)\cdot p\left(y_j\right)$, also wenn X und Y unabhängig sind. □

Lemma 3.12. *Sei (Ω,P) ein Wahrscheinlichkeitsraum, und seien X und Y auf Ω definierte Zufallsvariablen. Dann gilt $H(X\otimes Y)\geq H(X)$.*

Beweis: Sei $\{x_1,\dots,x_M\}$ der Bildbereich von X und $\{y_1,\dots,y_N\}$ der Bildbereich von Y. Es gilt

$$\begin{aligned}
H(X) &= \sum_{\substack{i=1\\ p(x_i)\neq 0}}^{M} -p(x_i)\log_2(p(x_i))\\
&= \sum_{\substack{i=1\\ p(x_i)\neq 0}}^{M} -\log_2(p(x_i))\cdot\sum_{j=1}^{N}p(x_i,y_j)\\
&= \sum_{\substack{(i,j)\in\{1,\dots,M\}\times\{1,\dots,N\}\\ p(x_i,y_j)>0}} -\log_2(p(x_i))\cdot p(x_i,y_j).
\end{aligned}$$

Da stets $p(x_i,y_j)\leq p(x_i)$ gilt, erhalten wir

$$\begin{aligned}
&\sum_{\substack{(i,j)\in\{1,\dots,M\}\times\{1,\dots,N\}\\ p(x_i,y_j)>0}} -\log_2(p(x_i))\cdot p(x_i,y_j)\\
&\leq \sum_{\substack{(i,j)\in\{1,\dots,M\}\times\{1,\dots,N\}\\ p(x_i,y_j)>0}} -\log_2(p(x_i,y_j))\cdot p(x_i,y_j)\\
&= H(X\otimes Y).
\end{aligned}$$

□

3. Zeichenweise Codierung

Wir versuchen nun, die Ausgänge eine Zufallsvariablen durch 0/1-Folgen, also durch Elemente in $\{0,1\}^+$ zu codieren. Manchmal kodieren wir nicht nur mit den zwei Symbolen 0 und 1, sondern allgemeiner mit Wörtern über einem D-elementigen Alphabet $\{b_1, b_2, \ldots, b_D\}$.

Definition 3.13. Sei B eine Menge und seien $c, d \in B^+$. Mit $c * d$ bezeichnen wir die Konkatenation von c und d.

Beispiel: Wenn $B = \{0,1\}$, $c = 01011$, $d = 10$, so gilt $c * d = 0101110$.

Definition 3.14 (Codierung). Sei A eine Menge, und sei $\{b_1, b_2, \ldots, b_D\}$ eine D-elementige Menge. Eine *Codierung* von A ist eine Funktion $C : A \to \{b_1, b_2, \ldots, b_D\}^+$. Die zu C gehörende *erweiterte Codierung* C^+ ist die Funktion

$$\begin{array}{rccc} C^+ & : & A^+ & \longrightarrow & \{b_1, b_2, \ldots, b_D\}^+ \\ & & (a_1, a_2, \ldots, a_n) & \longmapsto & C(a_1) * C(a_2) * \cdots * C(a_n). \end{array}$$

Definition 3.15. Sei A eine Menge, sei $\{b_1, b_2, \ldots, b_D\}$ eine D-elementige Menge, und sei C eine Codierung von A nach $\{b_1, b_2, \ldots, b_D\}^+$. C heißt *eindeutig decodierbar*, wenn C^+ injektiv ist.

Manchmal interessiert man sich nur für das Bild von A.

Definition 3.16. Sei $D \geq 2$, und sei $\{b_1, b_2, \ldots, b_D\}$ eine Menge. Eine Folge aus Wörtern in $\mathbf{C} = C_1, C_2, \ldots, C_M$ von $\{b_1, b_2, \ldots, b_D\}^+$ heißt auch *Code*. Der Code $\mathbf{C}$ ist *eindeutig decodierbar*, wenn al für alle $m, n \in \mathbb{N}$ und für alle $i_1, i_2, \ldots, i_m$ und $j_1, j_2, \ldots, j_n \in \{1, 2, \ldots, M\}$ folgendes gilt: Wenn

$$C_{i_1} * C_{i_2} * \cdots * C_{i_m} = C_{j_1} * C_{j_2} * \cdots * C_{j_n},$$

dann gilt $m = n$, und für alle $k \in \{1, \ldots, m\}$ gilt $i_k = j_k$.

Satz 3.17 (Ungleichung von Kraft und McMillan). *Sei $D \in \mathbb{N}$ mit $D \geq 2$, und sei $\mathbf{C} = (C_1, C_2, \ldots, C_M)$ ein eindeutig decodierbarer Code über dem Alphabet $\{a_1, a_2, \ldots, a_D\}$. Für $i \in \{1, 2, \ldots, M\}$ sei $n(i)$ die Länge von C_i. Dann gilt:*

$$\sum_{k=1}^{M} \frac{1}{D^{n(k)}} \leq 1.$$

Beweis: Für $m \in \mathbb{N}$ definieren wir $\mathbf{C}^m$ durch

$$\mathbf{C}^m := \{C_{i_1} * C_{i_2} * \cdots * C_{i_m} \mid i_1, i_2, \ldots, i_m \in \{1, 2, \ldots, M\}\}.$$

$\mathbf{C}^m$ besteht also aus allen Worten, die man durch Hintereinanderschreiben von genau m Codewörtern bilden kann. Wir berechnen nun $\left(\sum_{k=1}^{M} \frac{1}{D^{n(k)}}\right)^m$. Es gilt

$$\begin{aligned}\left(\sum_{k=1}^{M} \frac{1}{D^{n(k)}}\right)^m &= \sum_{(k_1,k_2,\ldots,k_m)\in\{1,2,\ldots,M\}^m} \frac{1}{D^{n(k_1)}} \cdot \frac{1}{D^{n(k_2)}} \cdot \ldots \cdot \frac{1}{D^{n(k_m)}} \\ &= \sum_{(k_1,k_2,\ldots,k_m)\in\{1,2,\ldots,M\}^m} \frac{1}{D^{n(k_1)+n(k_2)+\cdots+n(k_m)}}.\end{aligned} \tag{3.3}$$

Sei nun $x \in \mathbb{N}$. Wir überlegen uns nun, wie oft der Summand $\frac{1}{D^x}$ in dieser Summe auftritt. Dann gilt

$$\text{Anzahl der Summanden } \frac{1}{D^x} = \left|\left\{\mathbf{k} \in \{1,2,\ldots,M\}^m \;\middle|\; \sum_{i=1}^{m} n\left(k_i\right) = x\right\}\right|.$$

Wir betrachten nun die Abbildung φ, die durch

$$\begin{array}{rccl}\varphi & : & \{\mathbf{k} \in \{1,2,\ldots,M\}^m \mid \sum_{i=1}^{m} n\left(k_i\right) = x\} & \longrightarrow & \{W \in \mathbf{C}^m \mid \mathsf{L}(W) = x\} \\ & & (k_1, k_2, \ldots, k_m) & \longmapsto & C_{k_1} * C_{k_2} * \cdots * C_{k_m}\end{array}$$

definiert ist. Da der Code $\mathbf{C}$ eindeutig decodierbar ist, ist die Abbildung φ injektiv. Die Menge $\{W \in \mathbf{C}^m \mid \mathsf{L}(W) = x\}$ hat höchstens D^x Elemente (mehr Wörter der Länge x gibt es nicht). Die letzte Summe von (3.3) ist daher $\leq \sum_{x=1}^{\max\{n(i) \mid i\in\{1,2,\ldots,M\}\}\cdot m} D^x \cdot \frac{1}{D^x}$. Sei $N := \max\{n\left(i\right) \mid i \in \{1,2,\ldots,M\}\}$. Wir haben also insgesamt bewiesen, dass für alle $m \in \mathbb{N}$

$$\left(\sum_{k=1}^{M} \frac{1}{D^{n(k)}}\right)^m \leq N \cdot m$$

gilt. Es gilt also für alle $m \in \mathbb{N}$:

$$\sum_{k=1}^{M} \frac{1}{D^{n(k)}} \leq \sqrt[m]{N} \cdot \sqrt[m]{m}.$$

Da $\lim_{m\to\infty} \sqrt[m]{m} = \lim_{m\to\infty} e^{\frac{\log(m)}{m}} = e^{\lim_{m\to\infty} \frac{\log(m)}{m}} = e^0 = 1$, gilt also

$$\sum_{k=1}^{M} \frac{1}{D^{n(k)}} \leq 1.$$

□.

Satz 3.18 (Noiseless Coding Theorem – 1. Teil). *Sei X eine Zufallsvariable, sei $A = \{a_1, a_2, \ldots, a_M\}$ ihr Wertebereich. Für $i \in \{1,\ldots,M\}$ sei $p_i := P[X = a_i]$. Sei $D \in \mathbb{N}$ mit $D \geq 2$, und sei $C : \{a_1,\ldots,a_M\} \to \{b_1, b_2, \ldots, b_D\}^+$ eine eindeutig decodierbare Codierung. Seien $n_1, n_2, \ldots, n_M$ die Längen von $C\left(a_1\right), C\left(a_2\right), \ldots, C\left(a_M\right)$, und sei $\overline{n} := \sum_{i=1}^{M} p_i \cdot n_i$. Dann gilt $\overline{n} \geq \frac{H(X)}{\log_2(D)}$.*

Beweis: Lemma 3.9 liefert

$$H(p_1, p_2, \ldots, p_M) = \sum_{\substack{i=1 \\ p_i \neq 0}}^{M} -p_i \log_2(p_i) \leq \sum_{i=1}^{M} -p_i \log_2 \left(\frac{D^{-n_i}}{\sum_{j=1}^{M} D^{-n_j}} \right).$$

Also wissen wir

$$H(X) \leq \left(\sum_{i=1}^{M} -p_i \log_2 \left(D^{-n_i} \right) \right) + \log_2 \left(\sum_{j=1}^{M} D^{-n_j} \right) \cdot \left(\sum_{i=1}^{M} p_i \right).$$

Da wegen der Ungleichung von Kraft und McMillan (Satz 3.17) $\sum_{j=1}^{M} D^{-n_j} \leq 1$ gilt, ist der zweite Summand ≤ 0. Daher gilt

$$H(X) \leq \sum_{i=1}^{M} p_i \cdot n_i \cdot \log_2(D),$$

also

$$\frac{H(X)}{\log_2(D)} \leq \overline{n}.$$

□

4. Präfixfreie Codes

Definition 3.19. Sei $D \geq 2$, sei $\{b_1, b_2, \ldots, b_D\}$ eine Menge, und seien A, B Wörter aus $\{b_1, b_2, \ldots, b_D\}^+$. A ist ein *Präfix* von B, wenn $A = B$, oder wenn es ein Wort C gibt, sodass $A * C = B$.

Definition 3.20. Seien $A_1, A_2, \ldots, A_k$ Wörter über $\{b_1, b_2, \ldots, b_D\}$. Die Folge $(A_1, A_2, \ldots, A_k)$ ist ein *präfixfreier Code*, wenn es keine $i, j \in \{1, 2, \ldots, k\}$ gibt, sodass $i \neq j$ und A_i ein Präfix von A_j ist.

Satz 3.21. *Präfixfreie Codes sind eindeutig decodierbar.*

Beweis: Sei $\mathbf{C} = (C_1, C_2, \ldots, C_M)$ ein präfixfreier Code, seien $m, n \in \mathbb{N}$, und seien $i_1, i_2, \ldots, i_m$ und $j_1, j_2, \ldots, j_n \in \{1, \ldots, M\}$ so, dass

$$C_{i_1} * C_{i_2} * \cdots * C_{i_m} = C_{j_1} * C_{j_2} * \cdots * C_{j_n}. \tag{3.4}$$

Wir nehmen an, dass $m \leq n$. Falls für alle $k \in \{1, \ldots, m\}$ gilt: $i_k = j_k$, so muss wegen (3.4) $n = m$ gelten. Sei nun k minimal mit $i_k \neq j_k$. Es gilt nun $C_{i_k} * C_{i_{k+1}} * \cdots * C_{i_m} = C_{j_k} * C_{j_{k+1}} * \cdots * C_{j_n}$. Aufgrund dieser Gleichheit ist entweder C_{i_k} ein Präfix von C_{j_k} oder C_{j_k} ein Präfix von C_{i_k}. Da $i_k \neq j_k$ und $\mathbf{C}$ ein präfixfreier Code ist, ist das unmöglich. □

SATZ 3.22. *Sei $k \in \mathbb{N}$, sei $D \geq 2$, sei $\{b_1, b_2, \ldots, b_D\}$ eine Menge, und seien $n_1, n_2, \ldots, n_k \in \mathbb{N}$ mit $n_1 \leq n_2 \leq \cdots \leq n_k$. Dann gibt es genau dann einen präfixfreien Code $(A_1, A_2, \ldots, A_k)$ über $\{b_1, b_2, \ldots, b_D\}$, sodass $|A_1| = n_1$, ..., $|A_2| = n_2$, $|A_k| = n_k$, wenn*

$$\frac{1}{D^{n_1}} + \frac{1}{D^{n_2}} + \cdots + \frac{1}{D^{n_k}} \leq 1. \tag{3.5}$$

Beweis: Wir beweisen mit Induktion nach k, dass es einen präfixfreien Code mit den vorgegebenen Codewortlängen gibt, wenn die Ungleichung (3.5) erfüllt ist. Für $k = 1$ ist $A_1 := \underbrace{00\ldots0}_{n_1 \text{ mal}}$ das gesuchte Codewort. Sei nun $k \geq 2$. Da $\sum_{i=1}^{k} D^{-n_i} \leq 1$, gilt auch $\sum_{i=1}^{k-1} D^{-n_i} \leq 1$. Daher gibt es nach Induktionsvoraussetzung einen präfixfreien Code aus $k-1$ Wörtern $(A_1, \ldots, A_{k-1})$ mit Codewortlängen $(n_1, \ldots, n_{k-1})$. Als k-tes Wort dürfen wir jedes Wort wählen, das keines der A_i $(i = 1, \ldots, k-1)$ als Präfix hat. Von den D^{n_k} Wörtern der Länge n_k haben $D^{n_k - n_1}$ das Wort A_1 als Präfix, $D^{n_k - n_2}$ das Wort A_2, und $D^{n_k - n_{k-1}}$ das Wort A_k. Für die Auswahl des k-ten Wortes bleiben uns also zumindest $D^{n_k} - \sum_{i=1}^{k-1} D^{n_k - n_i}$ Wörter der Länge n_k übrig. Das ist wegen (3.5) echt größer als 0. □

KOROLLAR 3.23. *Sei $k \in \mathbb{N}$, sei $D \geq 2$, und seien $(n_1, n_2, \ldots, n_k)$ so, dass es es einen eindeutig decodierbaren Code $(A_1, A_2, \ldots, A_k)$ mit Codewortlängen $(n_1, n_2, \ldots, n_k)$ gibt. Dann gibt es sogar einen präfixfreien Code $(B_1, B_2, \ldots, B_k)$ mit Codewortlängen $(n_1, n_2, \ldots, n_k)$.*

Beweis: Wegen der Ungleichung von Kraft und McMillan (Satz 3.17) gilt

$$\sum_{i=1}^{k} D^{-n_i} \leq 1.$$

Satz 3.22 liefert dann einen präfixfreien Code mit den vorgegebenen Wortlängen. □

SATZ 3.24 (Noiseless Coding Theorem – 2. Teil). *Sei X eine Zufallsvariable mit Entropie $H(X)$. Dann gibt es eine präfixfreie Codierung C der Ausgänge von X auf $\{b_1, b_2, \ldots, b_D\}$, für deren durchschnittliche Codewortlänge $\overline{n} := E(\mathsf{L} \circ C \circ X)$ gilt*

$$\frac{H(X)}{\log_2(D)} \leq \overline{n} \leq \frac{H(X)}{\log_2(D)} + 1.$$

Beweis: Sei $M \in \mathbb{N}$ und seien $x_1, x_2, \ldots, x_M$ die Ausgänge von X. Für jedes $i \in \{1, \ldots, M\}$ sei $p_i := P[X = x_i]$. Wir nehmen an, dass $K \in \mathbb{N}$ so ist, dass $p_1 > 0, p_2 > 0, \ldots, p_K > 0$ und $p_{K+1} = \ldots = p_M = 0$. Wir wählen dann natürliche

Zahlen $n_1, \ldots, n_K$ so, dass

$$\log_D\left(\frac{1}{p_i}\right) \leq n_i \leq \log_D\left(\frac{1}{p_i}\right) + 1. \tag{3.6}$$

und dass für mindestens ein i die Ungleichung $\log_D\left(\frac{1}{p_i}\right) < n_i$ gilt. Da dann für $i \in \{1, \ldots, K\}$ die Ungleichung $-n_i \leq \log_D(p_i)$ (mit zumindest einmal $<$ statt $\leq$) gilt

$$\sum_{i=1}^{K} D^{-n_i} < \sum_{i=1}^{K} D^{\log_D(p_i)} = \sum_{i=1}^{K} p_i = 1.$$

Wir wählen nun $n_{K+1}, \ldots, n_M$ so, dass

$$n_j \geq \log_D\left(\frac{M-K}{1-\sum_{i=1}^{K} D^{-n_i}}\right) \text{ für } j \in \{K+1, \ldots, M\}.$$

Es gilt dann

$$\sum_{i=1}^{M} D^{-n_i} \leq 1.$$

Nach Satz 3.22 gibt es einen präfixfreien Code mit den Codewortlängen $n_1, n_2, \ldots, n_M$. Für seine durchschnittliche Wortlänge $\overline{n} = \sum_{i=1}^{K} p_i n_i$ erhalten wir aus Gleichung (3.6)

$$\sum_{i=1}^{K} p_i \cdot \log_D\left(\frac{1}{p_i}\right) \leq \sum_{i=1}^{K} p_i n_i \leq \sum_{i=1}^{K} p_i \cdot \left(\log_D\left(\frac{1}{p_i}\right) + 1\right).$$

Insgesamt erhalten wir

$$\frac{H(X)}{\log_2(D)} \leq \overline{n} \leq \frac{H(X)}{\log_2(D)} + 1.$$

□

5. Bedeutung der Entropie

Wir können jetzt die in der Sektion 1 definierten Größen $U_1(X)$ und $U_2(X)$ für den Informationsgehalt eines Ausgangs von X bestimmen.

Zuvor brauchen wir zwei Lemmas.

LEMMA 3.25. *Sei (Ω, P) ein Wahrscheinlichkeitsraum, sei $X : \Omega \to A$ eine Zufallsvariable, und sei $n \in \mathbb{N}$. Dann gilt $H(X^{[n]}) = n \cdot H(X)$.*

Beweis: Wir beweisen diesen Satz mit Induktion nach n. Für $n = 1$ ist die Behauptung unmittelbar klar. Für $n \geq 2$ definieren wir

$$X^{(n-1)}(\omega_1, \omega_2, \ldots, \omega_n) := (X(\omega_1), \ldots, X(\omega_{n-1})).$$

$X^{(n-1)}$ ist auf dem Wahrscheinlichkeitsraum Ω^n definiert. Diese Zufallsvariable nimmt die gleichen Werte, mit den gleichen Wahrscheinlichkeiten, an wie $X^{[n-1]}$. Daher gilt $H(X^{(n-1)}) = H(X^{[n-1]})$. Nun definieren wir

$$X^{(1)}(\omega_1, \omega_2, \ldots, \omega_n) := X(\omega_n).$$

$X^{(1)}$ nimmt die gleichen Werte, mit den gleichen Wahrscheinlichkeiten, an wie X. Daher gilt $H(X^{(1)}) = H(X)$. Die auf Ω^n definierte Zufallsvariable $X^{(n-1)} \otimes X^{(1)}$ liefert bis auf Umklammern die gleichen Ergebnisse wie die Funktion $X^{[n]}$. Daher gilt $H(X^{[n]}) = H(X^{(n-1)} \otimes X^{(1)})$. Da die Zufallsvariablen $X^{(n-1)}$ und $X^{(1)}$ unabhängig sind, gilt nach Korollar 3.11 $H(X^{(n-1)} \otimes X^{(1)}) = H(X^{(n-1)}) + H(X^{(1)}) = H(X^{[n-1]}) + H(X)$. Nach Induktionsvoraussetzung ist das $(n-1)H(X) + H(X) = nH(X)$. □

LEMMA 3.26. *Es gibt eine Präfixcodierung der natürlichen Zahlen C, sodass für alle $n \in \mathbb{N}$ für die Länge der Codierung $\mathsf{L}(C(n))$ gilt: $\mathsf{L}(C(n)) \leq 2\log_2(n) + 2$.*

Beweis: Sei $b(n)$ die Binärdarstellung von n. So ist zum Beispiel $b(8) = 1000$. Wir codieren nun die Zahl n so:

$$C(n) := \underbrace{11\ldots\ldots1}_{\mathsf{L}(b(n))-1 \text{ mal}} * 0 * b(n).$$

Es gilt $\mathsf{L}(C(n)) = 2 \cdot \lfloor \log_2(n) + 1 \rfloor$.

SATZ 3.27. *Sei (Ω, P) ein Wahrscheinlichkeitsraum, und sei $X : \Omega \to A$ eine Zufallsvariable. Dann gilt $U_1(X) = U_2(X) = H(X)$*

Dieser Satz folgt aus den folgenden drei Lemmas.

LEMMA 3.28. *Sei (Ω, P) ein Wahrscheinlichkeitsraum, und sei $X : \Omega \to A$ eine Zufallsvariable. Dann gilt $U_2(X) \leq U_1(X)$.*

Beweis: Wir zeigen, dass für alle $\varepsilon > 0$ gilt: $U_2(X) \leq U_1(X) + \varepsilon$. Aus der Definition von U_1 erhalten wir eine injektive Funktion $C : A^+ \to \{0,1\}^+$ mit $\liminf_{n\to\infty} \frac{1}{n} E(\mathsf{L} \circ C \circ X^{[n]}) \leq U_1(X) + \frac{\varepsilon}{2}$. Folglich gibt es eine unendliche Menge $M \subseteq \mathbb{N}$, sodass für alle $n \in M$ gilt: $\frac{1}{n} E(\mathsf{L} \circ C \circ X^{[n]}) \leq U_1(X) + \varepsilon$. Da die Einschränkung von C auf A^n injektiv ist, gilt für alle $n \in M$

$$\inf\{\frac{1}{n} E(\mathsf{L} \circ C' \circ X^{[n]}) \mid C' : A^n \to \{0,1\}^+, C' \text{ ist injektiv}\} \leq U_1(X) + \varepsilon.$$

Daher gilt auch $U_2(X) \leq U_1(X) + \varepsilon$. Somit haben wir $U_2(X) \leq U_1(X)$ bewiesen. □

LEMMA 3.29. *Sei (Ω, P) ein Wahrscheinlichkeitsraum, und sei $X : \Omega \to A$ eine Zufallsvariable. Dann gilt $U_1(X) \leq H(X)$.*

Wir zeigen, dass für alle $\varepsilon > 0$ gilt: $U_1(X) \leq H(X) + \varepsilon$. Wir wählen dazu $m \in \mathbb{N}$ so, dass $\frac{1}{m} \leq \frac{\varepsilon}{2}$. Wegen Satz 3.24 und Lemma 3.25 gibt es eine präfixfreie Codierung $F : A^m \to \{0,1\}^+$ der Ausgänge der Zufallsvariablen $X^{[m]}$, sodass

$$E(\mathsf{L} \circ F \circ X^{[m]}) \leq H(X^{[m]}) + 1.$$

Sei $M := \max\{\mathsf{L}(F(\mathbf{a})) \mid \mathbf{a} \in A^m\}$.

Wir codieren nun ein $\mathbf{a} \in A^+$ und definieren dazu Funktionen C_1, C_2, C_3 von A^+ nach $\{0,1\}^+$. Sei $n := \mathsf{L}(\mathbf{a})$. Es gilt also $\mathbf{a} = (a_1, a_2, \ldots, a_n)$.

- Als erstes codieren wir n von mit der Codierung aus Lemma 3.26. Sei c diese Codierung. Wir setzen
$$C_1(\mathbf{a}) := c(n).$$
- Nun codieren wir die ersten $\lfloor \frac{n}{m} \rfloor$ Blöcke von jeweils m Zeichen von $\mathbf{a}$ mithilfe von F. Es gilt also

$$C_2(\mathbf{a}) = F((a_1, \ldots, a_m)) * F((a_{m+1}, \ldots, a_{2m})) \\ * \cdots * F((a_{(\lfloor \frac{n}{m} \rfloor - 1)\cdot m + 1}, \ldots, a_{\lfloor \frac{n}{m} \rfloor \cdot m})).$$

- Es bleiben uns nun zwischen 0 und $m-1$ Zeichen übrig. Diese codieren wir mit C_3:
$$C_3(\mathbf{a}) := F((a_{\lfloor \frac{n}{m} \rfloor \cdot m + 1}, \ldots, a_{\lfloor \frac{n}{m} \rfloor \cdot m + (n - \lfloor \frac{n}{m} \rfloor \cdot m)}, \underbrace{a_n, \ldots, a_n}_{m - (n - \lfloor \frac{n}{m} \rfloor \cdot m) \text{ mal}})).$$

Als Codierung von $\mathbf{a}$ wählen wir nun $C(\mathbf{a}) := C_1(\mathbf{a}) * C_2(\mathbf{a}) * C_3(\mathbf{a})$. Die Funktion C ist eine injektive Funktion von A^+ nach $\{0,1\}^+$. Es gilt also $U_1(X) \leq \liminf_{n\to\infty} \frac{1}{n} E(\mathsf{L} \circ C \circ X^{[n]})$. Wir schätzen jetzt die Längen jedes einzelnen Teiles der Codierung ab: Es gilt

$$E(\mathsf{L} \circ C_1 \circ X^{[n]}) \leq 3 \log_2(n) \text{ für alle } n \geq 4.$$

Die Länge von C_2 kann so abgeschätzt werden.

$$E(\mathsf{L} \circ C_2 \circ X^{[n]}) = \lfloor \frac{n}{m} \rfloor \cdot E(\mathsf{L} \circ F \circ X^{[m]}) \\ \leq \frac{n}{m} \cdot (H(X^{[m]}) + 1) = \frac{n}{m} \cdot (m \cdot H(X) + 1).$$

Schließlich gilt

$$E(\mathsf{L} \circ C_3 \circ X^{[n]}) \leq M.$$

Insgesamt gilt

$$\frac{1}{n} \cdot E(\mathsf{L} \circ C \circ X^{[n]}) = \frac{1}{n} \cdot E(\mathsf{L} \circ C_1 \circ X^{[n]}) + \frac{1}{n} \cdot E(\mathsf{L} \circ C_2 \circ X^{[n]}) + \frac{1}{n} \cdot E(\mathsf{L} \circ C_3 \circ X^{[n]}) \\ \leq \frac{3 \log_2(n)}{n} + H(X) + \frac{1}{m} + \frac{M}{n}.$$

Für alle $n \in \mathbb{N}$ mit $n \geq 4$ und $\frac{3\log_2(n)}{n} + \frac{M}{n} \leq \frac{\varepsilon}{2}$ gilt also

$$\frac{1}{n} E(\mathsf{L} \circ C \circ X^{[n]}) \leq H(X) + \varepsilon.$$

Somit gilt $U_1(X) \leq H(X) + \varepsilon$. □

LEMMA 3.30. *Sei (Ω, P) ein Wahrscheinlichkeitsraum, und sei $X : \Omega \to A$ eine Zufallsvariable. Dann gilt $H(X) \leq U_2(X)$.*

Wir zeigen, dass für alle $\varepsilon > 0$ gilt: $H(X) \leq U_2(X) + \varepsilon$. Sei $\varepsilon > 0$. Es gibt unendlich viele n, für die es eine injektive Funktion $C : A^n \to \{0,1\}^+$ gibt, sodass $E(\mathsf{L} \circ C \circ X^{[n]}) \leq n \cdot (U_2(X) + \frac{\varepsilon}{2})$. Wir wählen unter diesen n ein N mit $N \geq 3$ und $\frac{3\log_2\left(N \cdot \log_2(|A|)+1\right)}{N} \leq \frac{\varepsilon}{2}$. Sei C eine injektive Funktion mit $E(\mathsf{L} \circ C \circ X^{[N]}) \leq N \cdot (U_2(X) + \frac{\varepsilon}{2})$.

Der Bildbereich von C hat $|A|^N$ Elemente. Für $M := \lceil N \cdot \log_2(|A|) \rceil$ gibt es also zumindest $2^{N \cdot \log_2(|A|)} = |A|^N$ 0/1-Folgen der Länge M. Wir können also aus C eine ebenfalls injektive Funktion C' bilden, sodass für alle $\mathbf{a} \in A^N$ gilt, dass $\mathsf{L}(C'(\mathbf{a})) \leq \mathsf{L}(C(\mathbf{a}))$ und $\mathsf{L}(C'(\mathbf{a})) \leq N \cdot \log_2(|A|) + 1$.

Wir bilden folgende eindeutig decodierbare Codierung der Ausgänge von $X^{[N]}$. Für $\mathbf{a} \in A^N$ bilden wir zunächst $C_1(\mathbf{a})$, indem wir die Länge von $C'(\mathbf{a})$ mit einem präfixfreien Code für alle natürlichen Zahlen, zum Beispiel mit dem Code c aus Lemma 3.26 codieren. Es gilt also

$$C_1(\mathbf{a}) = c(\mathsf{L}(C'(\mathbf{a}))).$$

Wir definieren nun

$$C_2(\mathbf{a}) := C_1(\mathbf{a}) * C'(\mathbf{a}).$$

Die Funktion C_2 ist injektiv. Sie ist zudem eine eindeutig decodierbare Codierung für die Ausgänge von $X^{[N]}$. (Das heißt, auch die erweiterte Codierung ${C_2}^+$ ist injektiv.) Es gilt daher wegen des Noiseless Coding Theorems (Satz 3.18), dass

$$E(\mathsf{L} \circ C_2 \circ X^{[N]}) \geq H(X^{[N]}).$$

Gleichzeitig gilt

$$\begin{aligned} E(\mathsf{L} \circ C_2 \circ X^{[N]}) &= E(\mathsf{L} \circ C_1 \circ X^{[N]}) + E(\mathsf{L} \circ C' \circ X^{[N]}) \\ &\leq 3\log_2\big(N \cdot \log_2(|A|) + 1\big) + N \cdot (U_2(X) + \frac{\varepsilon}{2}) \end{aligned}$$

Insgesamt gilt also

$$H(X) \leq \frac{3\log_2\big(N \cdot \log_2(|A|) + 1\big)}{N} + U_2(X) + \frac{\varepsilon}{2} \leq U_2(X) + \varepsilon.$$

Somit gilt $H(X) \leq U_2(X)$. □

6. Konstruktion optimaler Codes

DEFINITION 3.31. Sei $D \geq 2$, und seien $p_1, p_2, \ldots, p_M \in [0,1]$ so, dass $\sum_{i=1}^{M} p_i = 1$. Ein Code $\mathbf{C} = (C_1, C_2, \ldots, C_M)$ über dem Alphabet $\{b_1, b_2, \ldots, b_D\}$ ist *optimal* für $(p_1, p_2, \ldots, p_M)$, wenn er präfixfrei (und damit eindeutig decodierbar) ist, und wenn

$$\sum_{i=1}^{M} p_i \cdot \mathsf{L}(C_i) = \inf \{ \sum_{i=1}^{M} p_i \cdot \mathsf{L}(D_i) \mid (D_1, D_2, \ldots, D_M) \text{ ist präfixfreier Code über } \{b_1, b_2, \ldots, b_D\}\}.$$

Ein optimaler Code minimiert also die durchschnittliche Codewortlänge.

Das Infimum wird tatsächlich angenommen.

LEMMA 3.32. *Sei $D \geq 2$, und seien $p_1, p_2, \ldots, p_M \in [0,1]$ so, dass $\sum_{i=1}^{M} p_i = 1$ und $p_1 \geq p_2 \geq \cdots \geq p_M$. Es gibt einen optimalen Code für $(p_1, p_2, \ldots, p_M)$ über $\{b_1, b_2, \ldots, b_D\}$.*

Beweis: Wir betrachten zuerst den Fall, dass alle $p_i > 0$ sind. Eine Möglichkeit, die M Codewörter zu wählen, ist, alle M Wörter mit der Länge $\lceil \log_D(M) \rceil$ zu wählen. Dieser Code ist präfixfrei und hat durchschnittliche Codewortlänge $\lceil \log_D(M) \rceil$. Ein Code, bei dem irgendein Codewort länger als $\frac{\lceil \log_D(M) \rceil}{p_M}$ ist, hat sicher sicher eine höhere durchschnittliche Codewortlänge als $\lceil \log_D(M) \rceil$. Folglich gibt es nur endlich viele präfixfreie Codes, deren durchschnittliche Wortlänge $\leq \lceil \log_D(M) \rceil$ ist; unter diesen muss für einen die durchschnittliche Wortlänge minimal sein.

Wir betrachten nun den Fall, dass zumindest ein $p_i = 0$ ist. Seien $p_1, \ldots, p_K > 0$ und $p_{K+1} = \cdots = p_M = 0$. Es gibt einen präfixfreien Code mit durchschnittlicher Wortlänge $\lceil \log_2(M) \rceil$. Ein Code, für den eines der ersten K Codewörter länger als $\frac{\lceil \log_2(M) \rceil}{p_K}$ ist, hat durchschnittliche Wortänge $> \lceil \log_2(M) \rceil$. Wir wählen nun unter den (endlich vielen) präfixfreien Codes $(C_1, \ldots, C_K)$, für die $\sum_{i=1}^{K} D^{-\mathsf{L}(C_i)} < 1$ und für alle $i \in \{1, \ldots, K\}$ auch $\mathsf{L}(C_i) \leq \frac{\lceil \log_2(M) \rceil}{p_K}$ gilt, einen, der die durchschnittliche Codewortlänge für $(p_1, \ldots, p_K)$ minimiert. Sei $(D_1, \ldots, D_K)$ ein solcher Code. Da $\sum_{i=1}^{K} D^{-\mathsf{L}(C_i)} < 1$, können wir diesen Code zu einem präfixfreien Code

$$(D_1, D_2, \ldots, D_K, D_{K+1}, \ldots, D_M)$$

vervollständigen. Der so gewählte Code ist optimal: Nehmen wir an, ein präfixfreier Code $(E_1, E_2, \ldots, E_M)$ hat bessere durchschnittliche Codewortlänge. Dann gilt $\sum_{i=1}^{K} p_i \mathsf{L}(E_i) < \sum_{i=1}^{K} p_i \mathsf{L}(D_i)$. Wegen der Wahl der D_i muss daher $\sum_{i=1}^{K} D^{-\mathsf{L}(E_i)} = 1$ sein. Da $M > K$, verletzt E die Ungleichung von Kraft-McMillan; es kann ihn also nicht geben. □

Der Huffman-Algorithmus findet optimale Codes. Er beruht auf den folgenden Resultaten (cf. [**Ash90**, Übung 2.5, S. 301]).

LEMMA 3.33. *Seien $p_1, p_2, \dots, p_M \in [0,1]$ so, dass $\sum_{i=1}^{M} p_i = 1$, und sei $\mathbf{C} = (C_1, C_2, \dots, C_M)$ ein optimaler Code für $(p_1, p_2, \dots, p_M)$. Dann gilt für alle $i, j \in \{1, \dots, M\}$ mit $p_i > p_j$ auch $\mathsf{L}(C_i) \leq \mathsf{L}(C_j)$.*

Beweis: Wir bilden einen Code $(C'_1, C'_2, \dots, C'_M)$ durch $C'_k := C_k$ für $k \in \{1, 2, \dots, M\} \setminus \{i, j\}$, $C'_i := C_j$, $C'_j := C_i$. Da C optimal ist, gilt

$$\sum_{k=1}^{M} p_k \mathsf{L}(C_k) \leq \sum_{k=1}^{M} p_k \mathsf{L}(C'_k),$$

also

$$p_i \mathsf{L}(C_i) + p_j \mathsf{L}(C_j) \leq p_i \mathsf{L}(C_j) + p_j \mathsf{L}(C_i).$$

Das bedeutet $(p_i - p_j)(\mathsf{L}(C_j) - \mathsf{L}(C_i)) \geq 0$. Da $p_i - p_j > 0$, gilt daher $\mathsf{L}(C_i) \leq \mathsf{L}(C_j)$. □

SATZ 3.34. *Sei $D \geq 2$, sei $M \in \mathbb{N}$ mit $M > 1$ so, dass $M \equiv 1 \pmod{D-1}$, und seien $p_1, p_2, \dots, p_M \in [0,1]$ so, dass $\sum_{i=1}^{M} p_i = 1$ und $p_1 \geq p_2 \geq \cdots \geq p_M$. Dann gibt es einen optimalen Code für $(p_1, p_2, \dots, p_M)$ über $\{b_1, b_2, \dots, b_D\}$, sodass die D letzten Wörter $C_{M-D+1}, C_{M-D+2}, \dots, C_M$ die gleiche Länge haben (sagen wir, N) und in den ersten $N-1$ Zeichen übereinstimmen.*

Das heisst, dass es ein Wort Q der Länge $N-1$ und eine Bijektion $\pi : \{1, \dots, D\} \to \{1, \dots, D\}$ gibt, sodass gibt, sodass $C_{M-D+1} = Q * b_{\pi(1)}$, $C_{M-D+2} = Q * b_{\pi(2)}$, $\dots$, $C_M = Q * b_{\pi(D)}$.

Beweis: Wir starten mit einem optimalen Code $\mathbf{E}$, aus dem wir einen optimalen Code mit den gewünschten Eigenschaften bauen werden.

Zunächst wählen wir diesen optimalen Code $\mathbf{E}$ so unter allen optimalen Codes, dass $\sum_{i=1}^{M} \mathsf{L}(E_i)$ minimal ist.

Da $\mathbf{E}$ ein optimaler Code ist, gilt für $p_i > p_j$ auch $\mathsf{L}(E_i) \leq \mathsf{L}(E_j)$. Wir bilden nun den Code $\mathbf{F}$, indem wir in $\mathbf{E}$ die Wörter gleicher Wahrscheinlichkeit so ordnen, dass für $i \leq j$ immer $\mathsf{L}(F_i) \leq \mathsf{L}(F_j)$ gilt. Anschließend ordnen wir die Wörter gleicher Länge (unabhängig davon, ob sie gleich wahrscheinlich sind oder nicht) so um, dass die Wörter mit gleicher Länge lexikographisch aufsteigend (also wie im Telefonbuch) geordnet sind.

Der Code $\mathbf{F}$ hat die gleiche durchschnittliche Wortlänge wie $\mathbf{E}$, und die Codewortlängen sind schwach monoton wachsend. Sei $N := \mathsf{L}(F_M)$; N ist dann die maximale Codewortlänge von $\mathbf{F}$. Sei I die Anzahl der Codewörter von $\mathbf{F}$, die Länge N haben.

Wenn $I = 1$, so hat nur F_M die Länge N. Wir können dann das letzte Zeichen von F_M weglassen und erhalten einen präfixfreien Code, der ebenfalls optimal ist und bei dem die Summer der Codewortlängen um 1 kleiner als bei $\mathbf{E}$ ist. Das steht im Widerspruch zur Wahl von $\mathbf{E}$.

Wenn $I \geq D$, so spalten wir das letzte Codewort F_M auf. Seien Q ein Wort der Länge $N-1$ und $s \in \{1, 2, \ldots, D\}$ so, dass $F_M = Q * b_s$. Wir bilden einen neuen Code $\mathbf{G}$ durch $G_j := F_j$ für $j \leq M - D$, und $G_{M-D+j} := Q * b_j$. Auch $\mathbf{G}$ ist ein präfixfreier Code: kein Codewort der Länge $N-1$ kann Präfix eines $Q * b_j$ sein. Wenn ein Codewort G_k der Länge N mit $k \leq M - D$ gleich $Q * b_j$ wäre, dann können in $\mathbf{F}$ nach $G_k = F_k$ nur mehr die Wörter $Q * b_{j+1}, \ldots, Q * b_s$ stehen, da das die einzigen Wörter der Länge N sind, die lexikographisch zwischen $Q * b_j$ und $Q * b_s$ liegen. Dann hat $\mathbf{F}$ höchstens $M - D + D - 1$ Wörter, ein Widerspruch. Der Code $\mathbf{G}$ erfüllt also die gewünschten Eigenschaften.

Wenn $I \in \{2, 3, \ldots, D-1\}$, so wählen wir wieder das letzte Wort $F_M = Q * b_s$ und bilden einen Code $\mathbf{G}$ durch $G_j := F_j$ für $j \leq M - I$ und $G_{M-I+j} := Q * b_j$ für $j \in \{1, \ldots, I\}$. $\mathbf{G}$ ist ein präfixfreier Code. Nun zeigen wir:

> Jedes $R \in \{b_1, b_2, \ldots, b_D\}^{N-1} \setminus \{Q\}$ ist ein Codewort von $\mathbf{G}$, oder hat ein Codewort als (echten) Präfix.

Nehmen wir an R ist kein Codewort, $R \neq Q$, und R hat kein Codewort als Präfix. Nun bilden wir einen Code $\mathbf{H}$, indem wir in $\mathbf{G}$ das M-te Wort durch R ersetzen. Der so entstandene Code $\mathbf{H}$ ist ein präfixfreier Code und widerspricht der Wahl von $\mathbf{E}$. Daraus erhalten wir

$$D^{N-1} - 1 = \sum_{k=1}^{M-I} D^{N-1-\mathsf{L}(C_k)}.$$

Das ergibt modulo $D-1$:

$$0 \equiv M - I \pmod{D-1}.$$

Da $M \equiv 1 \pmod{D-1}$, gilt $I \equiv 1 \pmod{D-1}$. Das ist ein Widerspruch zu $I \in \{2, 3, \ldots, D-1\}$. Dieser Fall kann also nicht auftreten. □

Satz 3.35. *Sei $D \geq 2$, $M \equiv 1 \pmod{D-1}$, $M \geq D$. Seien $p_1, p_2, \ldots, p_M \in [0, 1]$ so, dass $\sum_{i=1}^{M} p_i = 1$ und $p_1 \geq p_2 \geq \cdots \geq p_M$. Wenn $\mathbf{C} = (C_1, C_2, \ldots, C_{M-D+1})$ ein optimaler Code für*

$$(p_1, p_2, \ldots, p_{M-D}, \sum_{i=1}^{D} p_{M-D+i})$$

ist, so ist

$$\mathbf{C}' := (C_1, C_2, \ldots, C_{M-D}, C_{M-D+1} * b_1, C_{M-D+1} * b_2, \ldots, C_{M-D+1} * b_D)$$

ein optimaler Code für $(p_1, p_2, \ldots, p_M)$.

Beweis: Sei $\mathbf{E}$ ein optimaler Code für $(p_1, p_2, \ldots, p_M)$, dessen letzte D Wörter die gleiche Länge haben und in allen Zeichen außer dem letzten übereinstimmen. Sei Q so, dass $E_M = Q * b_j$ mit $j \in \{1, \ldots, M\}$. Dann ist $(E_1, E_2, \ldots, E_{M-D}, Q)$ ein präfixfreier Code aus $M - D + 1$ Codewörtern. Denn wenn Q Präfix irgendeines E_i mit $i \in \{1, \ldots, M - D\}$ ist, so gibt es ein k mit $E_i = Q$ oder $E_i = Q * b_k$. Beides ist unmöglich. Da der Code $\mathbf{C}$ optimal für $(p_1, p_2, \ldots, p_{M-D+1})$ ist, gilt

(3.7)
$$\sum_{i=1}^{M-D} p_i \mathsf{L}(C_i) + \left(\sum_{i=M-D+1}^{M} p_i \right) \cdot \mathsf{L}(C_{M-D+1}) \leq \sum_{i=1}^{M-D} p_i \mathsf{L}(E_i) + \left(\sum_{i=M-D+1}^{M} p_i \right) \cdot (\mathsf{L}(E_M) - 1).$$

Nun berechnen wir die durchschnittliche Codewortlänge von $\mathbf{C}'$. Es gilt

Durchschnittliche Codewortlänge von $\mathbf{C}'$

$$= \sum_{i=1}^{M-D} p_i \mathsf{L}(C_i) + \sum_{i=M-D+1}^{M} p_i \cdot (\mathsf{L}(C_{M-D+1}) + 1)$$

$$= \sum_{i=1}^{M-D} p_i \mathsf{L}(C'_i) + \sum_{i=M-D+1}^{M} p_i \mathsf{L}(C_{M-D+1}) + \sum_{i=M-D+1}^{M} p_i.$$

Wegen (3.7) gilt

$$\sum_{i=1}^{M-D} p_i \mathsf{L}(C'_i) + \sum_{i=M-D+1}^{M} p_i \mathsf{L}(C_{M-D+1}) + \sum_{i=M-D+1}^{M} p_i$$

$$\leq \sum_{i=1}^{M-D} p_i \mathsf{L}(E_i) + \left(\sum_{i=M-D+1}^{M} p_i \right) \cdot (\mathsf{L}(E_M) - 1) + \sum_{i=M-D+1}^{M} p_i.$$

Da die letzten D Codewörter von $\mathbf{E}$ gleich lang sind, gilt

$$\sum_{i=1}^{M-D} p_i \mathsf{L}(E_i) + \left(\sum_{i=M-D+1}^{M} p_i \right) \cdot (\mathsf{L}(E_M) - 1) + \sum_{i=M-D+1}^{M} p_i$$

$$= \sum_{i=1}^{M-D} p_i \mathsf{L}(E_i) + \left(\sum_{i=M-D+1}^{M} p_i \cdot (\mathsf{L}(E_i) - 1) \right) + \sum_{i=M-D+1}^{M} p_i$$

$$= \sum_{i=1}^{M} p_i \mathsf{L}(E_i)$$

$$= \text{durchschnittliche Codewortlänge von } \mathbf{E}.$$

Da $\mathbf{E}$ optimal ist, ist also auch $\mathbf{C}'$ optimal. □

Dieser Satz liefert einen Algorithmus zur Konstruktion optimaler Codes, wenn $M \equiv 1 \pmod{D-1}$. Falls $M \not\equiv 1 \pmod{D-1}$, so erlaubt der folgende Satz,

Zeichen mit Wahrscheinlichkeit 0 hinzuzufügen, bis die Bedingung

$$M \equiv 1 \pmod{D-1}$$

erfüllt ist.

SATZ 3.36. *Sei $D \geq 2$, $M \geq 2$, und sei $r \in \{0, \ldots, D-2\}$ so, dass $M + r \equiv 1 \pmod{D-1}$. Seien $p_1, p_2, \ldots, p_M \in [0,1]$ so, dass $\sum_{i=1}^{M} p_i = 1$ und $p_1 \geq p_2 \geq \cdots \geq p_M$. Wenn $\mathbf{C} = (C_1, C_2, \ldots, C_{M+r})$ ein optimaler Code für*

$$(p_1, p_2, \ldots, p_M, \underbrace{0, 0, \ldots, 0}_{r \text{ mal}})$$

ist, so ist $(C_1, C_2, \ldots, C_M)$ ein optimaler Code für $(p_1, p_2, \ldots, p_M)$.

Beweis: Für $r = 0$ gilt der Satz offensichtlich. Sei nun $r \geq 1$. Wir nehmen an, der Code $(E_1, E_2, \ldots, E_M)$ erfüllt

$$\sum_{i=1}^{M} p_i \mathsf{L}(E_i) < \sum_{i=1}^{M} p_i \mathsf{L}(C_i).$$

Wenn $\sum_{i=1}^{M} D^{-\mathsf{L}(E_i)} < 1$, dann können wir $\mathbf{E}$ zu einem präfixfreien Code für $(p_1, p_2, \ldots, p_{M+r})$ erweitern. Dieser Code hat hat kürzere durchschnittliche Codewortlänge als $(C_1, C_2, \ldots, C_{M+r})$ und widerspricht somit dessen Optimalität. Also gilt $\sum_{i=1}^{M} D^{-\mathsf{L}(E_i)} = 1$. Sei N die maximale Codewortlänge von $\mathbf{E}$. Dann gilt

$$\sum_{i=1}^{M} D^{N-\mathsf{L}(E_i)} = D^N.$$

Modulo $D-1$ betrachtet bedeutet das

$$M \equiv 1 \pmod{D-1}.$$

Dann gilt $r = 0$. Das ist im Widerspruch zu $r \geq 1$ und zeigt, dass es den Code $\mathbf{E}$ mit kürzerer durchschnittlicher Codewortlänge als $(C_1, C_2, \ldots, C_M)$ nicht geben kann. Also ist $(C_1, C_2, \ldots, C_M)$ optimal. □

7. Eine andere Sichtweise des Quellcodierungssatzes

DEFINITION 3.37. In dieser Sektion wird ein Quellcodierungssatz ([**Mac03**, S. 78, Theorem 4.1], Satz 3.43) bewiesen, aus dem sich Teile des Noiseless Coding Theorems herleiten lassen.

Sei $X : \Omega \to A$ eine Zufallsvariable. Wir definieren die Zufallsvariable $\mathrm{LogP}(X)$ durch

$$\begin{array}{rccl} \mathrm{LogP}(X) & : & \Omega & \longrightarrow & \mathbb{R} \\ & & \omega & \longmapsto & \begin{cases} -\log_2(P[X = X(\omega)]) \text{ falls } P(\omega) > 0, \\ 0 \text{ sonst.} \end{cases} \end{array}$$

PROPOSITION 3.38. *Sei $X : \Omega \to A$ eine Zufallsvariable. Der Erwartungswert von $\mathrm{LogP}(X)$ ist $H(X)$.*

Beweis: Es gilt

$$\begin{aligned} E(\mathrm{LogP}(X)) &= \sum_{\substack{\omega\in\Omega\\ P(\omega)>0}} P(\omega)\cdot(-\log_2(P[X=X(\omega)])) \\ &= \sum_{\substack{\omega\in\Omega\\ P(\omega)>0}} P(\omega)\cdot \sum_{\substack{a\in A\\ P[X=a]>0}} (-\log_2(P[X=a]))\cdot\delta(X(\omega),a) \\ &= \sum_{\substack{a\in X\\ P[X=a]>0}} -\log_2(P[X=a])\cdot \sum_{\substack{\omega\in\Omega\\ P(\omega)>0}} P(\omega)\cdot\delta(X(\omega),a) \\ &= \sum_{\substack{a\in X\\ P[X=a]>0}} -\log_2(P[X=a])\cdot P[X=a] = H(X). \end{aligned}$$

□

DEFINITION 3.39. Sei $X : \Omega \to A$ eine Zufallsvariable, sei $n \in \mathbb{N}$, und sei $\beta > 0$. Wir nennen ein $\mathbf{a} \in A^n$ eine *typische Ausgangsfolge von X der Länge n zum Parameter β*, falls $P[X^{[n]} = \mathbf{a}] \neq 0$ und

$$\left| -\frac{1}{n}\log_2\left(P[X^{[n]}=\mathbf{a}]\right) - H(X)\right| < \beta.$$

Sei $T(X,n,\beta)$ die Menge der typischen Ausgangsfolgen von X der Länge n zum Parameter β.

Ein Ausgang $\mathbf{a} \in A^n$ ist also typisch, wenn

$$2^{-nH(X)+n\beta} \geq P[X^{[n]}=\mathbf{a}] \geq 2^{-nH(X)-n\beta}. \tag{3.8}$$

Auf lange Sicht haben wir es sehr wahrscheinlich mit typischen Ausgängen von X zu tun.

PROPOSITION 3.40. *Sei $X : \Omega \to A$ eine Zufallsvariable, sei $n \in \mathbb{N}$, und sei $\beta > 0$. Dann gilt*

$$P[X^{[n]} \in T(X,n,\beta)] \geq 1 - \frac{V(\mathrm{LogP}(X))}{\beta^2 n}. \tag{3.9}$$

Beweis: Das schwache Gesetz der großen Zahlen liefert für $\mathrm{LogP}(X)$:

$$P^{(n)}(\left\{\boldsymbol{\omega} \mid \left|\left(\frac{1}{n}\sum_{i=1}^{n}\mathrm{LogP}(X)(\omega_i)\right) - E(\mathrm{LogP}(X))\right| < \beta\right\}) \geq 1 - \frac{V(\mathrm{LogP}(X))}{\beta^2 n}.$$

Nun zeigen wir, dass für alle $\boldsymbol{\omega} \in \Omega^n$ mit $P^{(n)}(\boldsymbol{\omega}) > 0$ gilt:

$$\left| \left(\frac{1}{n} \sum_{i=1}^{n} \mathrm{LogP}(X)(\omega_i) \right) - E(\mathrm{LogP}(X)) \right| < \beta \text{ genau dann,} \tag{3.10}$$

wenn $X^{[n]}(\boldsymbol{\omega}) \in T(X, n, \beta)$.

Sei $\boldsymbol{\omega} \in \Omega^n$ so, dass $P^{(n)}(\boldsymbol{\omega}) > 0$. Wegen Proposition 3.38 gilt

$$\left| \left(\frac{1}{n} \sum_{i=1}^{n} \mathrm{LogP}(X)(\omega_i) \right) - E(\mathrm{LogP}(X)) \right| < \beta$$

genau dann, wenn

$$\left| -\frac{1}{n} \left(\sum_{i=1}^{n} \log_2(P[X = X(\omega_i)]) \right) - H(X) \right| < \beta.$$

Das ist äquivalent zu

$$\left| -\frac{1}{n} \log_2 \left(\prod_{i=1}^{n} P[X = X(\omega_i)] \right) - H(X) \right| < \beta,$$

also auch zu

$$\left| -\frac{1}{n} \log_2(P[X^{[n]} = X^{[n]}(\boldsymbol{\omega})]) - H(X) \right| < \beta.$$

Das Element $\boldsymbol{\omega}$ erfüllt diese Eigenschaft genau dann, wenn $X^{[n]}(\boldsymbol{\omega})$ in der Menge

$$\{\mathbf{a} \in A^n \mid \left| -\frac{1}{n} \log_2(P[X^{[n]} = \mathbf{a}]) - H(X) \right| < \beta\},$$

also in $T(X, n, \beta)$ liegt. Damit ist (3.10) gezeigt. Somit gilt auch (3.9). □

Wir wissen auch, wie viele Ausgänge ungefähr typisch sind.

PROPOSITION 3.41. *Sei $X : \Omega \to A$ eine Zufallsvariable, sei $n \in \mathbb{N}$, und sei $\beta > 0$. Dann gilt für die Anzahl der typischen Ausgänge von X:*

$$\left(1 - \frac{V(\mathrm{LogP}(X))}{\beta^2 n}\right) \cdot 2^{nH(X)-n\beta} \leq |T(X, n, \beta)| \leq 2^{nH(X)+n\beta}.$$

Beweis: Wir wissen:

$$P[X^{[n]} \in T(X, n, \beta)] = \sum_{\mathbf{a} \in T(X,n,\beta)} P[X^{[n]} = \mathbf{a}].$$

Wegen $P[X^{[n]} \in T(X, n, \beta)] \leq 1$ und (3.8) gilt

$$1 = \sum_{\mathbf{a} \in T(X,n,\beta)} P[X^{[n]} = \mathbf{a}] \geq \sum_{\mathbf{a} \in T(X,n,\beta)} 2^{-nH(X)-n\beta} = |T(X, n, \beta)| \cdot 2^{-nH(X)-n\beta}.$$

Daher gilt

$$|T(X,n,\beta)| \leq 2^{nH(X)+n\beta}.$$

Das beweist die zweite behauptete Ungleichung.

Um die erste Ungleichung zu beweisen, beobachten wir, dass wegen Proposition 3.40 gilt:

$$P[X \in T(X,n,\beta)] \geq 1 - \frac{V(\mathrm{LogP}(X))}{\beta^2 n}.$$

Also gilt

$$\sum_{\mathbf{a}\in T(X,n,\beta)} P[X^{[n]} = \mathbf{a}] \geq 1 - \frac{V(\mathrm{LogP}(X))}{\beta^2 n}.$$

Nach (3.8) gilt daher

$$|T(X,n,\beta)| \cdot 2^{-nH(X)+n\beta} \geq 1 - \frac{V(\mathrm{LogP}(X))}{\beta^2 n}$$

und somit

$$|T(X,n,\beta)| \leq \left(1 - \frac{V(\mathrm{LogP}(X))}{\beta^2 n}\right) \cdot 2^{nH(X)-n\beta}.$$

□

Definition 3.42. Sei $\delta \in (0,1)$, und sei $X : \Omega \to A$ eine Zufallsvariable. Wir definieren $H_\delta(X) := \log_2(\min\{|S| \mid S \subseteq A \text{ und } P[X \in S] \geq 1-\delta\})$.

Satz 3.43 (cf. [**Mac03**, Seite 78, Theorem 4.1]). *Sei X eine Zufallsvariable, und sei $\delta \in (0,1)$. Dann konvergiert die Folge $\langle \frac{1}{N} H_\delta(X^{[N]}) \mid N \in \mathbb{N}\rangle$ und es gilt $\lim_{N\to\infty} \frac{1}{N} H_\delta(X^{[N]}) = H(X)$.*

Beweis: Sei $\varepsilon > 0$.

Wir konstruieren als erstes n_0, sodass für alle $n \geq n_0$ gilt:

$$\frac{1}{n} H_\delta(X^{[n]}) \leq H(X) + \varepsilon.$$

Sei dazu n_0 so, dass $1 - \frac{V(\mathrm{LogP}(X))}{\varepsilon^2 \cdot n_0} \geq 1 - \delta$. Wegen Proposition 3.40 gilt dann für alle $n \geq n_0$:

$$P[X^{[n]} \in T(X,n,\varepsilon)] \geq 1 - \delta.$$

Daher gilt für alle $n \geq n_0$:

$$H_\delta(X^{[n]}) \leq \log_2(|T(X,n,\varepsilon)|).$$

Aus Proposition 3.41 folgt

$$H_\delta(X^{[n]}) \leq nH(X) + n\varepsilon.$$

Nun zeigen wir, dass es $n_0 \in \mathbb{N}$ gibt, sodass für alle $n \geq n_0$ gilt:

$$\frac{1}{n} H_\delta(X^{[n]}) \geq H(X) - \varepsilon. \tag{3.11}$$

Dazu zeigen wir, dass es ein n_0 gibt, sodass für alle $n \geq n_0$ und für alle Teilmengen S von A^n mit $|S| \leq 2^{nH(X)-n\varepsilon}$ gilt, dass $P[X^{[n]} \in S] < 1 - \delta$ ist. Für diese n gilt dann auch $H_\delta(X^{[n]}) \geq nH(X) - n\varepsilon$. Sei n_0 so, dass $2^{-n_0 \frac{\varepsilon}{2}} < \frac{1-\delta}{2}$ und $\frac{V(\mathrm{LogP}(X))}{(\frac{\varepsilon}{2})^2 \cdot n_0} < \frac{1-\delta}{2}$. Wir fixieren $n \in \mathbb{N}$ mit $n \geq n_0$. Es gilt nun

$$P[X^{[n]} \in S] \leq P\left[X^{[n]} \in S \cap T(X, n, \frac{\varepsilon}{2})\right] + P\left[X^{[n]} \notin T(X, n, \frac{\varepsilon}{2})\right].$$

Wegen (3.8) gilt

$$P\left[X^{[n]} \in S \cap T(X, n, \frac{\varepsilon}{2})\right] \leq |S| \cdot 2^{-nH(X)+n\frac{\varepsilon}{2}}.$$

Da $|S| \leq 2^{nH(X)-n\varepsilon}$, erhalten wir

$$P\left[X^{[n]} \in S \cap T(X, n, \frac{\varepsilon}{2})\right] \leq 2^{-n\frac{\varepsilon}{2}}.$$

Somit gilt

$$P\left[X^{[n]} \in S \cap T(X, n, \frac{\varepsilon}{2})\right] + P\left[X^{[n]} \notin T(X, n, \frac{\varepsilon}{2})\right] \leq 2^{-n\frac{\varepsilon}{2}} + \frac{V(\mathrm{LogP}(X))}{(\frac{\varepsilon}{2})^2 \cdot n}.$$

Da $n \geq n_0$, erhalten wir daraus

$$P[X^{[n]} \in S] < \frac{1-\delta}{2} + \frac{1-\delta}{2} = 1 - \delta.$$

Für $n \geq n_0$ gilt also auch (3.11). □

Aus Satz 3.43 lässt sich der 1. Teil des Noiseless Coding Theorems herleiten, und zwar ohne dass wir die Ungleichung von Kraft und McMillan brauchen.

2. Beweis von Satz 3.18: Sei $R := \sum_{i=1}^{M} p_i \cdot n_i$. Dann gilt also

$$E(\mathsf{L} \circ C \circ X) = R.$$

Wir zeigen, dass für alle $\varepsilon > 0$ gilt: $R \geq \frac{H(X)}{\log_2(D)} - \varepsilon$. Wir fixieren dazu $\varepsilon > 0$. Wir wählen $n_0 \in \mathbb{N}$ so, dass $\frac{V(\mathsf{L} \circ C \circ X)}{(\frac{\varepsilon}{2})^2 n_0} < \frac{1}{3}$. Sei nun $n \in \mathbb{N}$ mit $n \geq n_0$. Nach dem schwachen Gesetz der großen Zahlen gilt

$$P^{(n)}(\{\boldsymbol{\omega} \mid \left|\frac{1}{n}\sum_{i=1}^{n} \mathsf{L} \circ C \circ X(\omega_i) - R\right| < \frac{\varepsilon}{2}\}) \geq \frac{2}{3}.$$

Sei C^+ die erweiterte Codierung von $\{a_1, a_2, \ldots, a_M\}^+$ nach $\{b_1, b_2, \ldots, b_D\}^+$. Es gilt dann

$$\sum_{i=1}^{n} \mathsf{L} \circ C \circ X(\omega_i) = \mathsf{L} \circ C^+ \circ X^{[n]}(\boldsymbol{\omega}).$$

Es gilt also

$$P\left[\mathsf{L}\circ C^+\circ X^{[n]} < nR + n\frac{\varepsilon}{2}\right] \geq \frac{2}{3}. \tag{3.12}$$

Sei S die Menge, die durch

$$S := \{\mathbf{a} \in A^n \mid \mathsf{L}\circ C^+(\mathbf{a}) < nR + n\frac{\varepsilon}{2}\}$$

definiert wird. Da C eindeutig decodierbar ist, ist C^+ injektiv. Es gibt aber höchstens $\sum_{i=1}^{\lfloor nR+n\frac{\varepsilon}{2}\rfloor} D^i$ Wörter über einem D-elementigen Alphabet, deren Länge $< nR + n\frac{\varepsilon}{2}$ ist. Es gilt

$$\sum_{i=1}^{\lfloor nR+n\frac{\varepsilon}{2}\rfloor} D^i \leq \frac{D^{\lfloor nR+n\frac{\varepsilon}{2}\rfloor+1}-1}{D-1} \leq D^{nR+n\frac{\varepsilon}{2}+1}.$$

Wegen (3.12) gilt

$$P[X^{[n]} \in S] \geq \frac{2}{3}.$$

Für $\delta := \frac{1}{3}$ gilt also

$$H_\delta(X^{[n]}) \leq \log_2\left(D^{nR+n\frac{\varepsilon}{2}+1}\right).$$

Wir erhalten also

$$H_\delta(X^{[n]}) \leq \log_2(D)\cdot(nR + n\frac{\varepsilon}{2} + 1). \tag{3.13}$$

Da Gleichung (3.13) für alle $n > n_0$ gilt, gilt für alle $n > n_0$ mit $\frac{1}{n} \leq \frac{\varepsilon}{2}$ auch

$$H_\delta(X^{[n]}) \leq \log_2(D)\cdot(nR + n\varepsilon).$$

Nach Satz 3.43 gilt $\lim_{n\to\infty}\frac{1}{n}H_\delta(X^{[n]}) = H(X)$. Wir erhalten also

$$H(X) \leq \log_2(D)\cdot(R+\varepsilon),$$

und somit $\frac{H(X)}{\log_2(D)} - \varepsilon \leq R$. □

Wir können Satz 3.43 auch benutzen, um Codierungen zu konstruieren, die $H(X)$ nahe kommen.

KOROLLAR 3.44. *Sei $X : \Omega \to A$ eine Zufallsvariable mit Entropie $H(X)$, und sei $\varepsilon > 0$. Dann gibt es ein $n \in \mathbb{N}$ und eine präfixfreie Codierung C der Ausgänge von $X^{[n]}$ auf $\{b_1, b_2, \ldots, b_D\}$, für deren durchschnittliche Codewortlänge $E(\mathsf{L}\circ C\circ X^{[n]})$ gilt:*

$$\frac{1}{n}E(\mathsf{L}\circ C\circ X^{[n]}) \leq \frac{H(X)}{\log_2(D)} + \varepsilon.$$

Beweis: Wir wählen δ so, dass $\delta \leq \frac{\varepsilon}{3\log_D(|A|)}$. Aus Satz 3.43 erhalten wir ein $n_0 \in \mathbb{N}$, sodass für alle $n \geq n_0$ gilt: $H_\delta(X^{[n]}) \leq n \cdot (H(X) + \frac{\varepsilon}{3})$. Es gibt also für alle $n \in \mathbb{N}$ mit $n \geq n_0$ eine Menge $S_n \subseteq A^n$ mit $P[X \in S_n] \geq 1 - \delta$ und

$$|S_n| \leq 2^{nH(X)+\frac{n\varepsilon}{3}}.$$

Sei nun $n \in \mathbb{N}$ so, dass $n \geq n_0$ und $\frac{4}{n} \leq \frac{\varepsilon}{3}$. Die Elemente in S_n lassen sich also mit Wörtern, die alle die gleiche Länge $\lceil \log_D(|S_n|) \rceil$ haben, codieren. Sei C_1 diese Codierung.

Alle Folgen in A^n lassen sich mit Wörtern codieren, die alle die gleiche Länge $\lceil \log_D(|A|^n) \rceil$ haben. Sei C_2 diese Codierung.

Wir definieren nun eine Codierung $C : A^n \to \{b_1, b_2, \ldots, b_D\}^+$ folgendermaßen:

- Wenn $\mathbf{a} \in S_n$, so ist $C(\mathbf{a}) := 0 * C_1(\mathbf{a})$.
- Wenn $\mathbf{a} \notin S_n$, so ist $C(\mathbf{a}) := 1 * C_2(\mathbf{a})$.

C ist eine präfixfreie Codierung der Ausgänge von $X^{[n]}$. Wir bestimmen nun eine obere Schranke für $E(\mathsf{L} \circ C \circ X^{[n]})$. Es gilt

$$\begin{aligned}\frac{1}{n} \cdot E\left(\mathsf{L} \circ C \circ X^{[n]}\right) &\leq \frac{1}{n}\left(\left(1 + \log_D\left(2^{nH(X)+\frac{n\varepsilon}{3}}\right) + 1\right) + \delta \cdot \left(1 + \log_D\left(|A|^n\right) + 1\right)\right) \\ &\leq \frac{2}{n} + \log_D(2) \cdot H(X) + \frac{\log_D(2) \cdot \varepsilon}{3} + \frac{\delta \cdot 2}{n} + \delta \log_D(|A|) \\ &\leq \frac{4}{n} + \frac{H(X)}{\log_2(D)} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} \leq \frac{H(X)}{\log_2(D)} + \varepsilon.\end{aligned}$$

□

KAPITEL 4

Kanalcodierung

1. Bedingte Entropie

Seien X, Y Zufallsvariablen. Die *bedingte Entropie* $H(Y|X)$ gibt an, wieviele Bits wir zur Übertragung eines Ausganges von Y im Durchschnitt brauchen, wenn wir annehmen, dass dem Sender und dem Empfänger der Ausgang von X bekannt ist.

DEFINITION 4.1. Sei Ω ein Wahrscheinlichkeitsraum, und seien $X : \Omega \to A$ und $Y : \Omega \to B$ Zufallsvariablen. Dann definieren wir *die bedingte Entropie von Y, wenn X bekannt ist* durch $H(Y|X) := H(X \otimes Y) - H(X)$.

SATZ 4.2. *Sei Ω ein Wahrscheinlichkeitsraum, und seien $X : \Omega \to A$ und $Y : \Omega \to B$ Zufallsvariablen. Dann gilt*

$$H(Y|X) = \sum_{\substack{(a,b)\in A\times B \\ P[X=a \ \& \ Y=b]>0}} -P[X = a \ \& \ Y = b] \cdot \log_2(P[Y = b \mid X = a]).$$

Beweis: Es gilt

$$\sum_{\substack{(a,b)\in A\times B\\ P[X=a\ \&\ Y=b]>0}} -P[X=a\ \&\ Y=b]\cdot\log_2(P[Y=b \mid X=a])$$

$$= \sum_{\substack{(a,b)\in A\times B\\ P[X=a\ \&\ Y=b]>0}} -P[X=a\ \&\ Y=b]\cdot\log_2\left(\frac{P[Y=b\ \&\ X=a]}{P[X=a]}\right)$$

$$= \sum_{\substack{(a,b)\in A\times B\\ P[X=a\ \&\ Y=b]>0}} -P[X=a\ \&\ Y=b]\cdot\big(\log_2(P[Y=b\ \&\ X=a])-\log_2(P[X=a])\big)$$

$$= \sum_{\substack{(a,b)\in A\times B\\ P[X=a\ \&\ Y=b]>0}} -P[X=a\ \&\ Y=b]\cdot\log_2(P[Y=b\ \&\ X=a])$$

$$- \sum_{\substack{(a,b)\in A\times B\\ P[X=a\ \&\ Y=b]>0}} -P[X=a\ \&\ Y=b]\cdot\log_2(P[X=a])$$

$$= H(X\otimes Y) - \sum_{\substack{a\in A\\ P[X=a]>0}} -\log_2(P[X=a])\cdot \sum_{\substack{b\in B\\ P[X=a\ \&\ Y=b]>0}} P[X=a\ \&\ Y=b]$$

$$= H(X\otimes Y) - \sum_{\substack{a\in A\\ P[X=a]>0}} -\log_2(P[X=a])\cdot P[X=a]$$

$$= H(X\otimes Y) - H(X) = H(Y|X).$$

Definition 4.3. Sei Ω ein Wahrscheinlichkeitsraum, und sei $X:\Omega\to A$ eine Zufallsvariable. Für $a\in A$ mit $P[X=a]>0$ definieren wir eine Menge $\Omega|_{X=a}$ durch

$$\Omega_{|X=a} := \{\omega\in\Omega \mid X(\omega)=a\}.$$

Wir definieren auf dieser Menge ein Maß durch

$$P_{|X=a}(\omega) := \frac{P(\omega)}{P[X=a]}.$$

Proposition 4.4. *Sei Ω ein Wahrscheinlichkeitsraum, und seien $X:\Omega\to A$ und $Y:\Omega\to B$ Zufallsvariablen. Sei $a\in A$ mit $P[X=a]>0$. Dann ist die Funktion*

$$\begin{array}{rccc} Y|_{X=a} & : & \Omega_{|X=a} & \longrightarrow & B \\ & & \omega & \longmapsto & Y(\omega) \end{array}$$

eine Zufallsvariable auf dem Wahrscheinlichkeitsraum $(\Omega_{|X=a}, P_{|X=a})$. *Es gilt*

$$H(Y|X) = \sum_{\substack{a\in A\\ P[X=a]>0}} P[X=a]\cdot H(Y|_{X=a}).$$

Beweis:

$$\begin{aligned}
&\sum_{\substack{a\in A\\ P[X=a]>0}} P[X=a]\cdot H(Y|_{X=a})\\
=&\sum_{\substack{a\in A\\ P[X=a]>0}} P[X=a]\cdot \sum_{\substack{b\in B\\ P_{|X=a}[Y|_{X=a}=b]>0}} -P_{|X=a}[Y|_{X=a}=b]\cdot\log_2\big(P_{|X=a}[Y|_{X=a}=b]\big)\\
=&\sum_{\substack{a\in A\\ P[X=a]>0}} P[X=a]\cdot \sum_{\substack{b\in B\\ P[Y=b\ \&\ X=a]>0}} -\frac{P[Y=b\ \&\ X=a]}{P[X=a]}\cdot\log_2\big(P[Y=b \mid X=a]\big)\\
=&\sum_{a\in A}\ \sum_{\substack{b\in B\\ P[Y=b\ \&\ X=a]>0}} -P[Y=b\ \&\ X=a]\cdot\log_2\big(P[Y=b \mid X=a]\big)\\
&= H(Y|X).
\end{aligned}$$

Die letzte Gleichheit gilt wegen Satz 4.2.

□

Proposition 4.5. *Sei* Ω *ein Wahrscheinlichkeitsraum, und seien* $X:\Omega\to A$ *und* $Y:\Omega\to B$ *Zufallsvariablen. Wir definieren die Zufallsvariable* $\mathrm{LogP}(Y|X)$ *durch*

$$\begin{array}{rccl}\mathrm{LogP}(Y|X) & : & \Omega & \longrightarrow & \mathbb{R}\\ & & \omega & \longmapsto & \begin{cases} -\log_2(P[Y=Y(\omega) \mid X=X(\omega)]) \text{ falls } P(\omega)>0,\\ 0 \text{ sonst.}\end{cases}\end{array}$$

Dann gilt $E(\mathrm{LogP}(Y|X)) = H(Y|X)$.

Beweis: Wir verwenden in diesem Beweis wieder die Funktion δ mit der Eigenschaft $\delta(x,y)=1$, falls $x=y$ und $\delta(x,y)=0$, falls $x\neq y$.

$$
\begin{aligned}
E(\mathrm{LogP}(Y|X)) &= \sum_{\substack{\omega\in\Omega\\ P(\omega>0)}} P(\omega)\cdot(-\log_2(P[Y=Y(\omega)\mid X=X(\omega)]))\\
&= \sum_{\substack{\omega\in\Omega\\ P(\omega>0)}} P(\omega) \sum_{\substack{(a,b)\in A\times B\\ P[Y=b\ \&\ X=a]>0}} \big(-\log_2(P[Y=b\mid X=a])\big)\cdot\delta(Y(\omega),b)\cdot\delta(X(\omega),a))\\
&= \sum_{\substack{(a,b)\in A\times B\\ P[Y=b\ \&\ X=a]>0}} \big(-\log_2(P[Y=b\mid X=a])\big) \sum_{\substack{\omega\in\Omega\\ P(\omega)>0}} \delta(Y(\omega),b)\cdot\delta(X(\omega),a))\cdot P(\omega)\\
&= \sum_{\substack{(a,b)\in A\times B\\ P[Y=b\ \&\ X=a]>0}} \big(-\log_2(P[Y=b\mid X=a])\big)P[Y=b\ \&\ X=a] = H(Y|X).
\end{aligned}
$$

□

2. Eigenschaften der bedingten Entropie

Wir werden einige Eigenschaften der bedingten Entropie brauchen.

Satz 4.6. *Sei Ω ein Wahrscheinlichkeitsraum, und seien $X:\Omega\to A$ und $Y:\Omega\to B$ Zufallsvariablen. Dann gilt:*

(1) $0\leq H(Y|X)\leq H(Y)$.
(2) $H(Y\otimes Z|X)\leq H(Y|X)+H(Z|X)$.
(3) *(Kettenregel)* $H(Y\otimes Z|X)=H(Y|X)+H(Z|X\otimes Y)$.
(4) $H(Z|X\otimes Y)\leq H(Z|X)$.

Beweis: (1): Aus Satz 4.2 folgt $H(Y|X)\geq 0$, da alle Summanden nichtnegativ sind. Wegen Korollar 3.11 gilt $H(Y|X)=H(Y\otimes X)-H(X)\leq H(Y)+H(X)-H(X)=H(Y)$. Um (2) zu beweisen, benutzen wir die Darstellung der bedingten

Entropie aus Proposition 4.4 und erhalten

$$\begin{aligned}
H(Y \otimes Z|X) &= \sum_{\substack{a\in A\\ P[X=a]>0}} H((Y\otimes Z)|_{X=a}) \\
&= \sum_{\substack{a\in A\\ P[X=a]>0}} H(Y|_{X=a} \otimes Z|_{X=a}) \\
&\leq \sum_{\substack{a\in A\\ P[X=a]>0}} H(Y|_{X=a}) + H(Z|_{X=a}) \\
&= \sum_{\substack{a\in A\\ P[X=a]>0}} H(Y|_{X=a}) + \sum_{\substack{a\in A\\ P[X=a]>0}} H(Z|_{X=a}) \\
&= H(Y|X) + H(Z|X).
\end{aligned}$$

Für (3) berechnen wir $H(Y|X) + H(Z|X \otimes Y) = H(Y \otimes X) - H(X) + H(Z \otimes X \otimes Y) - H(X \otimes Y) = -H(X) + H(Z \otimes X \otimes Y) = H(Y \otimes Z|X)$. Um (4) zu beweisen, verwenden wir die Gleichung (3) und erhalten $H(Z|X \otimes Y) = H(Y \otimes Z|X) - H(Y|X)$. Nun verwenden wir (2) und erhalten $H(Y \otimes Z|X) - H(Y|X) \leq H(Y|X) + H(Z|X) - H(Y|X) = H(Z|X)$. □

Lemma 4.7. *Sei Ω ein Wahrscheinlichkeitsraum, sei $n \in \mathbb{N}$, und seien $X_1, X_2, \ldots, X_n$ und $Y_1, Y_2, \ldots, Y_n$ auf Ω definierte Zufallsvariablen. Dann gilt*

$$H(Y_1 \otimes \cdots \otimes Y_n|X_1 \otimes \cdots \otimes X_n) \leq \sum_{i=1}^{n} H(Y_i|X_i).$$

Beweis: Durch Induktion erhalten wir aus Satz 4.6 (2) die Ungleichung

$$H(Y_1 \otimes \cdots \otimes Y_n|X_1 \otimes \cdots \otimes X_n) \leq \sum_{i=1}^{n} H(Y_i|X_1 \otimes \cdots \otimes X_n).$$

Nun ist der i-te Summand dieser Summe wegen Satz 4.6 (4) höchstens $H(Y_i|X_i)$. Es gilt also $H(Y_1 \otimes \cdots \otimes Y_n|X_1 \otimes \cdots \otimes X_n) \leq \sum_{i=1}^{n} H(Y_i|X_i)$. □

Satz 4.8. *Sei Ω ein Wahrscheinlichkeitsraum, seien $n, M, N \in \mathbb{N}$, und seien $X_1, X_2, \ldots, X_n : \Omega \to \{1, 2, \ldots, M\}$ und $Y_1, Y_2, \ldots, Y_n : \Omega \to \{1, 2, \ldots, N\}$ Zufallsvariablen. Sei $\overline{X} := X_1 \otimes \cdots \otimes X_n$ und $\overline{Y} := Y_1 \otimes \cdots \otimes Y_n$. Wir nehmen an, dass für alle $(i_1, i_2, \ldots, i_n) \in \{1, 2, \ldots, M\}^n$ und $(j_1, j_2, \ldots, j_n) \in \{1, 2, \ldots, N\}^n$ mit $P[\overline{X} = (i_1, \ldots, i_n)] > 0$ gilt:*

$$P[\overline{Y} = (j_1, \ldots, j_n) \mid \overline{X} = (i_1, \ldots, i_n)] = \prod_{k=1}^{n} P[Y_k = j_k \mid X_k = i_k].$$

Dann gilt $H(\overline{Y}|\overline{X}) = \sum_{k=1}^{n} H(Y_k|X_k)$.

Beweis:

$$H(\overline{Y}|\overline{X}) =$$

$$= \sum_{\substack{\overline{j}\in\{1,2,\ldots,N\}^n \\ \overline{i}\in\{1,2,\ldots,M\}^n \\ P[\overline{Y}=\overline{j}\ \&\ \overline{X}=\overline{i}]>0}} -P[\overline{Y}=\overline{j}\ \&\ \overline{X}=\overline{i}]\log_2(P[\overline{Y}=\overline{j}\mid\overline{X}=\overline{i}])$$

$$= \sum_{k=1}^{n} \sum_{\substack{\overline{j}\in\{1,2,\ldots,N\}^n \\ \overline{i}\in\{1,2,\ldots,M\}^n \\ P[\overline{Y}=\overline{j}\ \&\ \overline{X}=\overline{i}]>0}} -P[\overline{Y}=\overline{j}\ \&\ \overline{X}=\overline{i}]\log_2(P[Y_k=j_k\mid X_k=i_k]).$$

Wir berechnen jetzt die innere Summe und erhalten

$$\sum_{\substack{\overline{j}\in\{1,2,\ldots,N\}^n \\ \overline{i}\in\{1,2,\ldots,M\}^n \\ P[\overline{Y}=\overline{j}\ \&\ \overline{X}=\overline{i}]>0}} - P[\overline{Y}=\overline{j}\ \&\ \overline{X}=\overline{i}]\log_2(P[Y_k=j_k\mid X_k=i_k])$$

$$= \sum_{\substack{\overline{j}\in\{1,2,\ldots,N\}^n \\ \overline{i}\in\{1,2,\ldots,M\}^n \\ P[\overline{Y}=\overline{j}\ \&\ \overline{X}=\overline{i}]>0}} \sum_{\substack{(a,b)\in\{1,2,\ldots,M\}\times\{1,2,\ldots,N\} \\ P[Y_k=b\ \&\ X_k=a]>0}} -P[\overline{Y}=\overline{j}\ \&\ \overline{X}=\overline{i}]\cdot$$

$$\log_2(P[Y_k=b\mid X_k=a])\cdot\delta(j_k,b)\cdot\delta(i_k,a)$$

$$= \sum_{\substack{(a,b)\in\{1,2,\ldots,M\}\times\{1,2,\ldots,N\} \\ P[Y_k=b\ \&\ X_k=a]>0}} -\log_2(P[Y_k=b\mid X_k=a])\cdot$$

$$\sum_{\substack{\overline{j}\in\{1,2,\ldots,N\}^n \\ \overline{i}\in\{1,2,\ldots,M\}^n}} \delta(j_k,b)\cdot\delta(i_k,a)\cdot P[\overline{Y}=\overline{j}\ \&\ \overline{X}=\overline{i}]$$

$$= \sum_{\substack{(a,b)\in\{1,2,\ldots,M\}\times\{1,2,\ldots,N\} \\ P[Y_k=b\ \&\ X_k=a]>0}} -\log_2(P[Y_k=b\mid X_k=a])\cdot P[Y_k=b\ \&\ X_k=a]$$

$$= H(Y_k|X_k).$$

□

3. Bedeutung der bedingten Entropie

Wir geben nun eine Interpretation von $H(Y|X)$ an. Für $\mathbf{a} \in A^n$ und $\mathbf{b} \in B^n$ schreiben wir $(\mathbf{a}, \mathbf{b})^T$ für den Vektor $((a_1, b_1), (a_2, b_2), \ldots, (a_n, b_n))$ in $(A \times B)^n$.

Definition 4.9. Sei Ω ein Wahrscheinlichkeitsraum, seien $X : \Omega \to A$ und $Y : \Omega \to B$ Zufallsvariablen, und sei $n \in \mathbb{N}$. Eine Funktion $C : (A \times B)^n \to \{0,1\}^+$ ist *eine gültige Codierung von n Ausgängen von Y auf der Basis von X*, wenn für alle $\mathbf{a} \in A^n$ die Abbildung

$$\begin{array}{rccl} C_{\mathbf{a}} & : & B^n & \longrightarrow & \{0,1\}^+ \\ & & \mathbf{b} & \longmapsto & C((\mathbf{a}, \mathbf{b})^T) \end{array}$$

injektiv ist. Wir definieren $U_2(Y|X)$, den Informationsgehalt von Y auf der Basis von X, durch

$$U_2(Y|X) := \liminf_{n\to\infty} \inf\Big\{ \frac{1}{n} E(\mathsf{L} \circ C \circ (X \otimes Y)^{[n]}) \;\Big|\; C \text{ ist eine gültige Codierung von } n \text{ Ausgängen von } Y \text{ auf der Basis von } X \Big\}.$$

Satz 4.10. *Sei Ω ein Wahrscheinlichkeitsraum, und seien $X : \Omega \to A$ und $Y : \Omega \to B$ Zufallsvariablen. Dann gilt $U_2(Y|X) = H(Y|X)$.*

Wir zeigen als erstes $U_2(Y|X) \leq H(Y|X)$. Sei $n \in \mathbb{N}$. Wir betrachten die Zufallsvariablen $X^{[n]}$ und $Y^{[n]}$, die beide auf dem gleichen Wahrscheinlichkeitsraum Ω^n definiert sind. Sei $\mathbf{a} \in A^n$ mit $P[X^{[n]} = \mathbf{a}] > 0$. Wir betrachten die auf $\Omega^n_{|X^{[n]}=\mathbf{a}}$ definierte Zufallsvariable $Y^{[n]}|_{X^{[n]}=\mathbf{a}}$. Wir finden eine präfixfreie Codierung $C_{\mathbf{a}}$ der Ausgänge dieser Zufallsvariablen, die im Durchschnitt höchstens $H(Y^{[n]}|_{X^{[n]}=\mathbf{a}}) + 1$ Bits pro Ausgang braucht. Das bedeutet

$$\sum_{\mathbf{b}\in B^n} \frac{P[Y^{[n]} = \mathbf{b} \;\&\; X^{[n]} = \mathbf{a}]}{P[X^{[n]} = \mathbf{a}]} \cdot \mathsf{L}(C_{\mathbf{a}}(\mathbf{b})) \leq H(Y^{[n]}|_{X^{[n]}=\mathbf{a}}) + 1. \tag{4.1}$$

Für die $\mathbf{a} \in A^n$ mit $P[X^{(n)} = \mathbf{a}] = 0$ wählen wir als $C_{\mathbf{a}}$ irgendeine injektive Funktion von B^n nach $\{0,1\}^+$. Wir definieren nun $C : (A \times B)^n \to \{0,1\}^+$ durch

$$C((\mathbf{a}, \mathbf{b})^T) := C_{\mathbf{a}}(\mathbf{b}).$$

Wir berechnen nun den Erwartungswert von $\mathsf{L} \circ C \circ (X \otimes Y)^{[n]}$. Es gilt

$$E(\mathsf{L} \circ C \circ (X \otimes Y)^{[n]} = \sum_{\mathbf{a} \in A^n} \sum_{\mathbf{b} \in B^n} P[X^{[n]} = \mathbf{a} \;\&\; Y^{[n]} = \mathbf{b}] \cdot \mathsf{L}(C((\mathbf{a}, \mathbf{b})^T))$$

$$= \sum_{\substack{\mathbf{a} \in A^n \\ P[X^{[n]} = \mathbf{a}] > 0}} P[X^{[n]} = \mathbf{a}] \cdot \sum_{\mathbf{b} \in B^n} \frac{P[Y^{[n]} = \mathbf{b} \;\&\; X^{[n]} = \mathbf{a}]}{P[X^{[n]} = \mathbf{a}]} \cdot \mathsf{L}(C_{\mathbf{a}}(\mathbf{b}))$$

$$\leq \sum_{\substack{\mathbf{a} \in A^n \\ P[X^{[n]} = \mathbf{a}] > 0}} P[X^{[n]} = \mathbf{a}] \cdot \left(H(Y^{[n]}|_{X^{[n]} = \mathbf{a}}) + 1\right)$$

$$= H(Y^{[n]}|X^{[n]}) + 1.$$

Wir verwenden nun Lemma 4.7 für die Zufallsvariablen $X_i : \Omega^n \to A, \boldsymbol{\omega} \mapsto X(\omega_i)$ und $Y_i : \Omega^n \to B, \boldsymbol{\omega} \mapsto Y(\omega_i)$ und erhalten

$$H(Y^{[n]}|X^{[n]}) + 1 \leq \sum_{k=1}^{n} H(Y_k|X_k) + 1 = n \cdot H(Y|X) + 1.$$

Somit erhalten wir $U_2(Y|X) \leq \liminf_{n \to \infty} H(Y|X) + \frac{1}{n}$, und somit $U_2(Y|X) \leq H(Y|X)$.

Nun zeigen wir $U_2(Y|X) \geq H(Y|X)$. Dazu zeigen wir, dass für alle $\varepsilon > 0$ gilt: $U_2(Y|X) \geq H(Y|X) - \varepsilon$. Sei dazu $\varepsilon > 0$. Für alle $n \in \mathbb{N}$ gibt es eine präfixfreie Codierung C_X von $X^{[n]}$, die im Durschnitt weniger als $H(X^{[n]}) + 1$ Bits pro Ausgang von $X^{[n]}$ braucht. Wegen der Definition von $U_2(Y|X)$ gibt es für unendlich viele $n \in \mathbb{N}$ eine gültige Codierung C von n Ausgaïgen von Y auf der Basis von X mit

$$E(\mathsf{L} \circ C \circ (X \otimes Y)^{[n]}) \leq n \cdot (U_2(Y|X) + \frac{\varepsilon}{3}).$$

Da für jedes $\mathbf{a} \in A^n$ die zugehörige Funktion $C_{\mathbf{a}}$ nur $|B|^n$ verschiedene Werte annehmen muss, können wir jedes $C_{\mathbf{a}}$ durch eine injektive Funktion $C'_{\mathbf{a}}$ ersetzen, sodass für alle $\mathbf{b} \in B^n$ gilt $\mathsf{L}(C'_{\mathbf{a}}(\mathbf{b})) \leq \mathsf{L}(C_{\mathbf{a}}(\mathbf{b}))$ und $\mathsf{L}(C'_{\mathbf{a}}(\mathbf{b})) \leq \log_2(|B|^n) + 1$. Wir wählen jetzt unter diesen unendlich vielen $n \in \mathbb{N}$ eines mit $\frac{2\log_2(n \log_2(|B|)+1)+2}{n} \leq \frac{\varepsilon}{3}$ und $\frac{1}{n} \leq \frac{\varepsilon}{3}$. Wir bilden dann folgende eindeutig decodierbare Codierung der Ausgänge von $(X \otimes Y)^{[n]}$.

Sei $c : \mathbb{N} \to \{0,1\}^+$ eine präfixfreie Codierung der natürlichen Zahlen, die für alle $n \in \mathbb{N}$ die Abschätzung

$$\mathsf{L}(c(n)) \leq 2\log_2(n) + 2$$

erfüllt. Für $\mathbf{a} \in A^n$ und $\mathbf{b} \in B^n$ codieren wir $(\mathbf{a}, \mathbf{b})^T$ durch

$$C_{X \otimes Y}((\mathbf{a}, \mathbf{b})^T) := C_X(\mathbf{a}) * c(\mathsf{L}(C'_{\mathbf{a}})(\mathbf{b})) * C'_{\mathbf{a}}(\mathbf{b}).$$

Der Erwartungswert für die Länge dieser Codierung $C_{X\otimes Y}$ lässt sich durch

$$E(\mathsf{L}\circ C_{X\otimes Y}\circ (X\otimes Y)^{[n]}) \leq (H(X^{[n]})+1)+(2\log_2(n\log_2(|B|)+1)+2)+(n\cdot(U_2(Y|X)+\frac{\varepsilon}{3}))$$

abschätzen. Wegen des Noiseless Coding Theorems muss also gelten

$$(H(X^{[n]})+1)+(2\log_2(n\log_2(|B|)+1)+2)+(n\cdot(U_2(Y|X)+\frac{\varepsilon}{3}))\geq H((X\otimes Y)^{[n]}),$$

also

$$(nH(X)+1)+(2\log_2(n\log_2(|B|)+1)+2)+(n\cdot(U_2(Y|X)+\frac{\varepsilon}{3}))\geq n\,H(X\otimes Y).$$

Wir divideren diese Ungleichung durch n und erhalten

$$H(X)+\frac{\varepsilon}{3}+\frac{\varepsilon}{3}+U_2(Y|X)+\frac{\varepsilon}{3}\geq H(X\otimes Y),$$

also

$$U_2(X)\geq H(Y|X)-\varepsilon.$$

□

4. Der gegenseitige Informationsgehalt

Definition 4.11. Sei Ω ein Wahrscheinlichkeitsraum, und seien $X:\Omega\to A$ und $Y:\Omega\to B$ Zufallsvariablen. Die gegenseitige Information zwischen X und Y ist definiert durch

$$I(X;Y):=H(X)+H(Y)-H(X\otimes Y).$$

Es gilt $I(X;Y)=H(X)-H(X|Y)$. Die Größe $I(X;Y)$ gibt also an, wieviele Bits man sich bei der Übertragung von X spart, wenn man Y kennt. Ebenso gilt $I(X;Y)=H(Y)-H(Y|X)$. $I(X;Y)$ gibt also zugleich an, wieviele Bits man sich bei der Übertragung von Y spart, wenn man X kennt.

5. Markovketten

Definition 4.12. Sei Ω ein Wahrscheinlichkeitsraum, sei $n\in\mathbb{N}$, seien $A_1,A_2,\ldots,A_n$ Mengen, und sei für $i\in\{1,2,\ldots,n\}$ die Funktion $X_i:\Omega\to A_i$ eine Zufallsvariable. Die Folge $(X_1,X_2,\ldots,X_n)$ ist eine *Markovkette*, wenn für alle $j\in\{2,\ldots,n\}$ und für alle $(a_1,a_2,\ldots,a_j)\in\prod_{i=1}^{j}A_i$ mit

$$P[X_1=a_1\ \&\ X_2=a_2\ \&\ \ldots\ \&\ X_{j-1}=a_{j-1}]>0$$

gilt:

$$P[X_j=a_j\mid X_{j-1}=a_{j-1}] = P[X_j=a_j\mid X_{j-1}=a_{j-1}\ \&\ X_{j-2}=a_{j-2}\ \&\ \ldots\ \&\ X_1=a_1]. \tag{4.2}$$

LEMMA 4.13. *Sei Ω ein Wahrscheinlichkeitsraum, sei $n \in \mathbb{N}$, seien $A_1, A_2, \ldots, A_n$ Mengen, und sei für $i \in \{1, 2, \ldots, n\}$ die Funktion $X_i : \Omega \to A_i$ eine Zufallsvariable. Wir nehmen an, dass $(X_1, X_2, \ldots, X_n)$ eine Markovkette ist.*

(1) *Seien $r \in \mathbb{N}_0$, $i_1, i_2, \ldots, i_r \in \mathbb{N}$ und $k \in \{1, \ldots, n-1\}$ so, dass*

$$i_1 < \cdots < i_r < k.$$

Dann gilt für alle $\mathbf{a} \in \prod_{j=1}^{k=1} A_j$ mit $P[X_k = a_k \ \& \ X_{i_1} = a_{i_1} \ \& \ \ldots X_{i_r} = a_{i_r}] > 0$:

$$P[X_{k+1} = a_{k+1} \mid X_k = a_k \ \& \ X_{i_1} = a_{i_1} \ \& \ \ldots X_{i_r} = a_{i_r}] = P[X_{k+1} = a_{k+1} \mid X_k = a_k].$$

(2) *Seien $i < j < k$. Dann gilt für alle $(a_i, a_j, a_k) \in A_i \times A_j \times A_k$ mit $P[X_k = a_k \ \& \ X_j = a_j \ \& \ X_i = a_i] > 0$:*

$$P[X_k = a_k \mid X_j = a_j \ \& \ X_i = a_i] = P[X_k = a_k \mid X_j = a_j].$$

Beweis: Wir beweisen (1) durch Induktion nach $k - r$. Falls $k - r = 1$, dann ist die Behauptung genau die Markov-Eigenschaft (4.2). Wenn $k - r \geq 2$, dann wählen wir ein $j \in \{1, 2, \ldots, k-1\} \setminus \{i_1, \ldots, i_r\}$. Wir kürzen nun ab:

$$p(a_{j_1}, \ldots, a_{j_s}) := P[X_{j_1} = a_{j_1} \ \& \ \cdots \ \& \ X_{j_s} = a_{j_s}].$$

Sei $(a_{i_1}, \ldots, a_{i_r}, a_k)$ so, dass $p(a_{i_1}, \ldots, a_{i_r}, a_k) > 0$. Dann gilt

$$p(a_{i_1}, \ldots, a_{i_r}, a_k, a_{k+1}) = \sum_{a_j \in A_j} p(a_{i_1}, \ldots, a_{i_r}, a_j, a_k, a_{k+1}).$$

Wir nehmen nun an, dass a_j so ist, dass $p(a_{i_1}, \ldots, a_{i_r}, a_j, a_k, a_{k+1}) \neq 0$. Dann gilt

$$p(a_{i_1}, \ldots, a_{i_r}, a_j, a_k, a_{k+1}) = p(a_{i_1}, \ldots, a_{i_r}, a_j, a_k) \cdot p(a_{k+1} \mid a_{i_1}, \ldots, a_{i_r}, a_j, a_k).$$

Nach Induktionsvoraussetzung gilt also

$$p(a_{i_1}, \ldots, a_{i_r}, a_j, a_k, a_{k+1}) = p(a_{i_1}, \ldots, a_{i_r}, a_j, a_k) \cdot p(a_{k+1} \mid a_k).$$

Somit gilt

$$\begin{aligned} &\sum_{a_j \in A_j} p(a_{i_1}, \ldots, a_{i_r}, a_j, a_k, a_{k+1}) \\ &\qquad = \sum_{a_j \in A_j} p(a_{i_1}, \ldots, a_{i_r}, a_j, a_k) \cdot p(a_{k+1} \mid a_k) \\ &\qquad = p(a_{k+1} \mid a_k) \cdot p(a_{i_1}, \ldots, a_{i_r}, a_k). \end{aligned}$$

Daraus erhalten wir

$$\frac{p(a_{i_1}, \ldots, a_{i_r}, a_k, a_{k+1})}{p(a_{i_1}, \ldots, a_{i_r}, a_k)} = p(a_{k+1} \mid a_k).$$

Das beweist (1).

Wir beweisen nun (2) durch Induktion nach $k-j$. Falls $k-j=1$, so ist die Gleichung (2) ein Spezialfall der bereits bewiesenen Gleichung (1). Falls $k-j \geq 2$, so gilt (mit den gleichen Abkürzungen wie im Beweis von (1)):

$$p(a_j, a_k) = \sum_{\substack{a_{k-1} \in A_{k-1} \\ p(a_j, a_{k-1})>0}} p(a_j, a_{k-1}, a_k)$$

$$= \sum_{\substack{a_{k-1} \in A_{k-1} \\ p(a_j, a_{k-1})>0}} p(a_j) \cdot p(a_{k-1} \mid a_j) \cdot p(a_k \mid a_j, a_{k-1}).$$

Nach Induktionsvoraussetzung gilt $p(a_{k-1} \mid a_j) = p(a_{k-1} \mid a_i, a_j)$. Gemeinsam mit (1) erhalten wir

$$\sum_{\substack{a_{k-1} \in A_{k-1} \\ p(a_j, a_{k-1})>0}} p(a_j) \cdot p(a_{k-1} \mid a_j) \cdot p(a_k \mid a_j, a_{k-1})$$

$$= \sum_{\substack{a_{k-1} \in A_{k-1} \\ p(a_j, a_{k-1})>0}} p(a_j) \cdot p(a_{k-1} \mid a_i, a_j) \cdot p(a_k \mid a_i, a_j, a_{k-1})$$

$$= p(a_j) \cdot \sum_{\substack{a_{k-1} \in A_{k-1} \\ p(a_j, a_{k-1})>0}} p(a_{k-1}, a_k \mid a_i, a_j)$$

$$= p(a_j) \cdot p(a_k \mid a_i, a_j)$$

Somit gilt $p(a_k \mid a_j) = p(a_k \mid a_i, a_j)$. □

Aus (2) dieses Lemmas folgt sofort:

KOROLLAR 4.14. *Sei $n \geq 3$, sei $(X_1, X_2, \ldots, X_n)$ eine Markovkette, und seien $i, j, k \in \mathbb{N}$ so, dass $i < j < k \leq n$. Dann ist (X_i, X_j, X_k) eine Markovkette.*

LEMMA 4.15. *Sei (X, Y, Z) eine Markovkette. Dann gilt:*

(1) $H(Z|X \otimes Y) = H(Z|Y)$;
(2) $H(X|Y \otimes Z) = H(X|Y)$.

Beweis: Für alle x, y, z bilden wir folgende Abkürzungen: $p(x, y, z) := P[X = x \;\&\; Y = y \;\&\; Z = z]$, $p(z \mid x) := P[Z = z \mid X = x]$. Wir beweisen nun (1). Nach

Satz 4.2 gilt

$$\begin{aligned} H(Z|X\otimes Y) &= \sum_{\substack{(x,y,z)\in A\times B\times C\\ p(x,y,z)>0}} -p(x,y,z)\cdot\log_2(p(z\mid x,y)) \\ &= \sum_{\substack{(x,y,z)\in A\times B\times C\\ p(x,y,z)>0}} -p(x,y,z)\cdot\log_2 p(z\mid y) \\ &= \sum_{\substack{(y,z)\in B\times C\\ p(y,z)>0}} -p(y,z)\log_2(p(z\mid y)) = H(Z|Y). \end{aligned}$$

Für (2) zeigen wir zunächst, dass mit (X,Y,Z) auch (Z,Y,X) eine Markovkette ist. Tatsächlich gilt, falls $p(x,y,z)>0$, auch

$$p(x\mid y,z) = \frac{p(x,y,z)}{p(y,z)} = \frac{p(x,y)\cdot p(z\mid x,y)}{p(y)\cdot p(z\mid y)}.$$

Wir verwenden nun, dass (X,Y,Z) eine Markovkette ist, und erhalten

$$\frac{p(x,y)\cdot p(z\mid x,y)}{p(y)\cdot p(z\mid y)} = \frac{p(x,y)\cdot p(z\mid y)}{p(y)\cdot p(z\mid y)} = p(x\mid y).$$

Somit ist auch (Z,Y,X) eine Markovkette. Nach (1) gilt also $H(X|Y) = H(X|Y\otimes Z)$. □

Lemma 4.16. *Sei (X,Y,Z) eine Markovkette. Dann gilt $I(X;Z)\le I(Y;Z)$ und $I(X;Z)\le I(X;Y)$.*

Beweis: Es gilt $I(Y;Z) = H(Z) - H(Z|Y)$. Wegen Lemma 4.15 gilt $H(Z) - H(Z|Y) = H(Z) - H(Z|X\otimes Y)$. Da $H(Z|X\otimes Y)\le H(Z|X)$, gilt $H(Z) - H(Z|X\otimes Y)\ge H(Z) - H(Z|X) = I(X;Z)$. Für den zweiten Teil berechnen wir $I(X;Y) = H(X) - H(X|Y)$. Wegen Lemma 4.15 gilt $H(X) - H(X|Y) = H(X) - H(X|Y\otimes Z)\ge H(X) - H(X|Z) = I(X;Z)$. □

Satz 4.17. *Sei $(X_1,\dots,X_n)$ eine Markovkette, und sei $1\le i<j\le n$. Dann gilt $I(X_i;X_j)\ge I(X_1;X_n)$.*

Beweis: Da (X_1,X_i,X_n) eine Markovkette ist, gilt $I(X_1;X_n)\le I(X_i;X_n)$. Da (X_i,X_j,X_n) eine Markovkette ist, gilt $I(X_i;X_n)\le I(X_i;X_j)$. □

6. Kanäle

Definition 4.18. Seien $M,N\in\mathbb{N}$. Eine *stochastische $M\times N$-Matrix* ist eine $M\times N$-Matrix mit Einträgen aus $[0,1]$, deren Zeilensummen alle gleich 1 sind.

Seien $X : \Omega \to \{x_1, x_2, \ldots, x_M\}$ und $Y : \Omega \to \{y_1, y_2, \ldots, y_N\}$ Zufallsvariablen, und sei $P[X = x_i] > 0$ für alle $i \in \{1, \ldots, M\}$. Dann ist die Matrix A mit $A(i,j) := P[Y = y_j \mid X = x_i]$ eine stochastische $M \times N$-Matrix.

DEFINITION 4.19. Seien $M, N \in \mathbb{N}$. Ein *Kanal* ist ein Tripel

$$\big((x_1, x_2, \ldots, x_M), (y_1, y_2, \ldots, y_N), A\big),$$

wobei A eine stochastische $M \times N$-Matrix ist, die Zeichen $x_1, x_2, \ldots, x_M$ paarweise verschieden sind, und die Zeichen $y_1, y_2, \ldots, y_N$ ebenfalls paarweise verschieden sind. Die Zeichen $x_1, \ldots, x_M$ heißen *Eingabezeichen*, die Zeichen $y_1, \ldots, y_N$ heißen *Ausgabezeichen* des Kanals.

DEFINITION 4.20. Sei $\mathbf{K} = \big((x_1, x_2, \ldots, x_M), (y_1, y_2, \ldots, y_N), A\big)$ ein Kanal, und sei $X : \Omega \to \{x_1, x_2, \ldots, x_M\}$ eine Zufallsvariable. Eine Zufallsvariable $Y : \Omega \to \{y_1, y_2, \ldots, y_N\}$ ist *eine zum Kanal* $\mathbf{K}$ *und zur Eingabe* X *passende Ausgabe*, wenn für alle $i \in \{1, \ldots, M\}$ mit $P[X = x_i] > 0$ gilt: $A(i,j) = P[Y = y_j \mid X = x_i]$.

SATZ 4.21. *Sei* $\mathbf{K} = \big((x_1, x_2, \ldots, x_M), (y_1, y_2, \ldots, y_N), A\big)$ *ein Kanal, und seien* X, Y *Zufallsvariablen. Wir nehmen an, dass* Y *eine zu* $\mathbf{K}$ *und* X *passende Ausgabe ist. Für* $i \in \{1, \ldots, M\}$ *sei* $p_i := P[X = x_i]$ *und* A_i *der* i*-te Zeilenvektor von* A*. Dann gilt*

$$I(X;Y) = H((p_1, p_2, \ldots, p_M) \cdot A) - \sum_{i=1}^{M} p_i \cdot H(A_i).$$

Beweis: Wir berechnen $I(X;Y)$ als $H(Y) - H(Y|X)$. Zunächst sehen wir, dass für alle $j \in \{1, \ldots, N\}$ gilt:

$$\begin{aligned} P[Y = y_j] &= \sum_{i=1}^{M} P[X = x_i \;\&\; Y = y_j] \\ &= \sum_{i=1}^{M} P[X = x_i] \cdot A(i,j) = \sum_{i=1}^{M} p_i A(i,j). \end{aligned} \tag{4.3}$$

Folglich gilt $H(Y) = H((p_1, p_2, \ldots, p_M) \cdot A)$.

Nun berechnen wir $H(Y|X)$. Es gilt

$$\begin{aligned} &H(Y|X) \\ &\quad= \sum_{\substack{(i,j)\in\{1,\dots,M\}\times\{1,\dots,N\} \\ P[X=x_i \ \&\ Y=y_j]>0}} P[X=x_i \ \&\ Y=y_j]\cdot\log_2(P[Y=y_j \mid X=x_i]) \\ &\qquad= \sum_{\substack{(i,j)\in\{1,\dots,M\}\times\{1,\dots,N\} \\ P[X=x_i \ \&\ Y=y_j]>0}} p_i\cdot A(i,j)\cdot\log_2(A(i,j)) \\ &\qquad= \sum_{\substack{i=1 \\ p_i>0}}^{M}\ \sum_{\substack{j=1 \\ A(i,j)>0}}^{N} p_i\cdot A(i,j)\cdot\log_2(A(i,j)) \\ &\qquad\qquad= \sum_{\substack{i=1 \\ p_i>0}}^{M} p_i H(A_i) = \sum_{i=1}^{M} p_i H(A_i). \end{aligned}$$

□

Definition 4.22. Sei A eine stochastische $M\times N$-Matrix, und seien $p_1, p_2, \dots, p_M \in [0,1]$ so, dass $\sum_{i=1}^{M} p_i = 1$. Für $i \in \{1,\dots,M\}$ sei A_i der i-te Zeilenvektor von A. Dann definieren wir

$$T\big(A,(p_1,p_2,\dots,p_M)\big) := H((p_1,p_2,\dots,p_M)\cdot A) - \sum_{i=1}^{M} p_i\cdot H(A_i).$$

Wir nennen $T\big(A,(p_1,p_2,\dots,p_M)\big)$ den *Transinformationsgehalt von A bezüglich der Wahrscheinlichkeiten $(p_1,p_2,\dots,p_M)$*.

Satz 4.23. *Sei $\mathbf{K}$ ein Kanal mit Matrix A, und seien X, Y Zufallsvariablen, sodass Y eine zu $\mathbf{K}$ und X passende Ausgabe ist. Sei $p_i := P[X=x_i]$. Dann gilt $I(X;Y) = T(A,(p_1,p_2,\dots,p_M))$.*

Definition 4.24. Sei $\mathbf{K}$ ein Kanal mit M Eingabezeichen und Kanalmatrix A. Dann ist die *Kanalkapazität* C von $\mathbf{K}$ durch

$$C := \sup\ \{T(A,(p_1,p_2,\dots,p_M)) \mid p_1,p_2,\dots,p_M \in [0,1], \sum_{i=1}^{M} p_i = 1\}$$

definiert

Für eine zum Kanal $\mathbf{K}$ und zur Eingabe X gehörende Ausgabe Y gilt also wegen Satz 4.21 immer $I(X;Y) \le C$.

Wir verwenden jetzt den Kanal mehrmals.

DEFINITION 4.25. Sei $n \in \mathbb{N}$, sei $\mathbf{K} = ((x_1, x_2, \ldots, x_M), (y_1, y_2, \ldots, y_N), A)$ ein Kanal, und sei $\overline{X} : \Omega \to \{x_1, x_2, \ldots, x_M\}^n$, $\overline{Y} : \Omega \to \{y_1, y_2, \ldots, y_N\}^n$. Dann ist $\overline{Y}$ eine zur n-facher Verwendung von $\mathbf{K}$ und zur Eingabefolge $\overline{X}$ passende Ausgabefolge, wenn für alle $(x_{i_1}, x_{i_2}, \ldots, x_{i_n})$ mit $P[\overline{X} = (x_{i_1}, x_{i_2}, \ldots, x_{i_n})] > 0$ gilt: $P[\overline{Y} = (y_{j_1}, y_{j_2}, \ldots, y_{j_n}) \mid \overline{X} = (x_{i_1}, x_{i_2}, \ldots, x_{i_n})] = \prod_{k=1}^{n} A(i_k, j_k)$.

LEMMA 4.26. *Sei $n \in \mathbb{N}$, sei $\mathbf{K} = ((x_1, x_2, \ldots, x_M), (y_1, y_2, \ldots, y_N), A)$ ein Kanal, und sei $\overline{X} : \Omega \to \{x_1, x_2, \ldots, x_M\}^n$, $\overline{Y} : \Omega \to \{y_1, y_2, \ldots, y_N\}^n$, und sei $\overline{Y}$ eine zur n-facher Verwendung von $\mathbf{K}$ und zur Eingabefolge $\overline{X}$ passende Ausgabefolge. Sei $k \in \{1, \ldots, n\}$, sei X_k die k-te Komponente von $\overline{X}$, und sei Y_k die k-te Komponente von $\overline{Y}$. Dann gilt für alle $i_k \in \{1, \ldots, M\}$ mit $P[X_k = x_{i_k}] > 0$ und alle $j_k \in \{1, 2, \ldots, N\}$:*

$$P[Y_k = y_{j_k} \mid X_k = x_{i_k}] = A(i_k, j_k).$$

Beweis: Wir führen den Beweis für $k = 1$. Sei i so, dass $P[X_1 = x_{i_1}] > 0$. Wir kürzen $P[X_1 = x_{i_1} \;\&\; \cdots \;\&\; X_n = x_{i_n} \;\&\; Y_1 = y_{j_1} \;\&\; \cdots \;\&\; Y_n = y_{j_n}]$ mit $p(i_1, \ldots, i_n, j_1, \ldots j_n)$ ab und erhalten:

$$\begin{aligned}
&P[X_1 = x_{i_1} \;\&\; Y_1 = y_{j_1}] \\
&\quad = p(i_1, j_1) \\
&\quad = \sum_{\substack{(i_2,\ldots,i_n)\in\{1,\ldots,M\}^{n-1} \\ (j_2,\ldots,j_n)\in\{1,\ldots,N\}^{n-1}}} p(i_1, \ldots, i_n, j_1, \ldots, j_n) \\
&\quad = \sum_{\substack{(i_2,\ldots,i_n)\in\{1,\ldots,M\}^{n-1} \\ (j_2,\ldots,j_n)\in\{1,\ldots,N\}^{n-1}}} p(i_1, \ldots, i_n) \cdot A(i_1, j_1) \cdot \prod_{k=2}^{n} A(i_k, j_k) \\
&\quad = \sum_{(i_2,\ldots,i_n)\in\{1,\ldots,M\}^{n-1}} p(i_1, \ldots, i_n) \cdot A(i_1, j_1) \cdot \sum_{(j_2,\ldots,j_n)\in\{1,\ldots,N\}^{n-1}} \prod_{k=2}^{n} A(i_k, j_k) \\
&\quad = \sum_{(i_2,\ldots,i_n)\in\{1,\ldots,M\}^{n-1}} p(i_1, \ldots, i_n) \cdot A(i_1, j_1) \cdot \prod_{k=2}^{n} \Big(\sum_{j\in\{1,\ldots,N\}} A(i_k, j) \Big) \\
&\quad = \sum_{(i_2,\ldots,i_n)\in\{1,\ldots,M\}^{n-1}} p(i_1, \ldots, i_n) \cdot A(i_1, j_1) \cdot \prod_{k=2}^{n} 1 \\
&\quad = p(i_1) \cdot A(i_1, j_1).
\end{aligned}$$

SATZ 4.27. *Sei* $\mathbf{K}$ *ein Kanal mit Kanalkapazität* C*, sei* $n \in \mathbb{N}$*, und sei* $\overline{Y}$ *eine zur* n*-facher Verwendung von* $\mathbf{K}$ *und zur Eingabefolge* $\overline{X}$ *passende Ausgabefolge. Dann gilt* $I(\overline{X};\overline{Y}) \leq n \cdot C$.

Beweis: Für $k \in \{1,\ldots,n\}$ sei X_k die k-te Komponente von $\overline{X}$ und Y_k die k-te Komponente von $\overline{Y}$. Wegen Lemma 4.26 gilt, wenn $P[X_k = x_{i_k}] > 0$:

$$P[Y_k = y_{j_k} \mid X_k = x_{i_k}] = A(i_k, j_k). \tag{4.4}$$

Somit ist die Voraussetzung von Satz 4.8 erfüllt, und es gilt $H(\overline{Y}|\overline{X}) = \sum_{k=1}^{n} H(Y_k|X_k)$. Wegen (4.4) ist Y_k eine zur Eingabe X_k und zum Kanal $\mathbf{K}$ passende Ausgabe. Somit gilt $I(X_k;Y_k) \leq C$. Es gilt nun

$$\begin{aligned} I(\overline{X};\overline{Y}) &= H(\overline{Y}) - H(\overline{Y}|\overline{X}) \\ &\leq \sum_{k=1}^{n} H(Y_k) - H(\overline{Y}|\overline{X}) \\ &= \sum_{k=1}^{n} H(Y_k) - \sum_{k=1}^{n} H(Y_k|X_k) \\ &= \sum_{k=1}^{n} H(Y_k) - H(Y_k|X_k) \\ &= \sum_{k=1}^{n} I(Y_k;X_k) \\ &\leq n \cdot C. \end{aligned}$$

□

7. Untere Schranken für den Übertragungsfehler

DEFINITION 4.28. Sei $\mathbf{K}$ ein Kanal mit Eingabezeichen $(x_1, x_2, \ldots, x_M)$, Ausgabezeichen $(y_1, y_2, \ldots, y_N)$ und Kanalmatrix A. Seien $U, \overline{X}, \overline{Y}, V$ auf Ω definierte Zufallsvariablen. (Wir stellen uns vor, dass $U(\omega)$ eine Nachricht, $\overline{X}(\omega)$ eine Codierung dieser Nachricht für den Kanal, $\overline{Y}(\omega)$ die Ausgabe aus dem Kanal, und $V(\omega)$ die decodierte Nachricht ist.) Wir nennen $(U, \overline{X}, \overline{Y}, V)$ ein *Übertragungssystem für binäre Nachrichten der Länge* m *mit* n*-facher Benutzung des Kanals* $\mathbf{K}$, falls folgende Eigenschaften erfüllt sind:

(1) $(U, \overline{X}, \overline{Y}, V)$ ist eine Markovkette,
(2) $U : \Omega \to \{0,1\}^m$,
(3) $\overline{X} : \Omega \to \{x_1, x_2, \ldots, x_M\}^n$,
(4) $\overline{Y} : \Omega \to \{y_1, y_2, \ldots, y_N\}^n$,

(5) $\overline{Y}$ ist eine zur n-fachen Verwendung von $\mathbf{K}$ und zur Eingabefolge $\overline{X}$ passende Ausgabefolge,
(6) $V : \Omega \to \{0,1\}^m$.

Die *Übertragungsrate* des Systems ist gegeben durch $R := \frac{m}{n}$ [Nachrichtenbits pro Kanalzeichen].

DEFINITION 4.29. Sei $(U, \overline{X}, \overline{Y}, V)$ ein auf Ω definiertes Übertragungssystem für binäre Nachrichten der Länge m. Dann definieren wir folgende Größen, um den Übertragungsfehler zu messen:

(1) den i-ten Einzelfehler F_i, der bestimmt, ob die i-te Komponente der empfangenen Nachricht mit der i-ten Komponente der gesendeten Nachricht übereinstimmt. Es gilt $F_i(\omega) = 1$, falls $U(\omega)(i) \neq V(\omega)(i)$, und $F_i(\omega) = 0$ sonst.
(2) die mittlere Bitfehlerrate p_B, die durch $p_B := \frac{1}{m} E(\sum_{i=1}^m F_i)$ definiert ist.
(3) die Blockfehlerrate p_e, die durch $p_e := P[\sum_{i=1}^n F_i \geq 1]$ definiert ist.

SATZ 4.30. *Sei* $\mathbf{K}$ *ein Kanal mit Kanalkapazität* C*, und sei* $(U, \overline{X}, \overline{Y}, V)$ *ein auf* Ω *definiertes Übertragungssystem für binäre Nachrichten der Länge* m*. Wir nehmen an, dass alle Nachrichten gleichwahrscheinlich sind, also dass* $P[U = \mathbf{x}] = 2^{-m}$ *für alle* $\mathbf{x} \in \{0,1\}^m$*. Sei* p_e *die Blockfehlerrate und* p_B *die Bitfehlerrate des Übertragungssystems. Es gilt:*

(1) $p_e \geq 1 - \frac{C}{R} - \frac{1}{m}$.
(2) $H(p_B, 1 - p_B) \geq 1 - \frac{C}{R}$.
(3) $p_e \geq p_B \geq \frac{p_e}{m}$.

Beweis: Für den Beweis von (1) definieren wir eine Funktion F auf Ω durch $F(\omega) = 1$, falls $\sum_{i=1}^m F_i(\omega) \geq 1$, und $F(\omega) = 0$, falls $\sum_{i=1}^m F_i(\omega) = 0$. Für den Erwartungswert $E(F)$ gilt dann $E(F) = p_e$. Wegen der Kettenregel gilt

$$H(U \otimes F|V) = H(U|V) + H(F|V \otimes U).$$

Ebenso gilt

$$H(U \otimes F|V) = H(F|V) + H(U|V \otimes F).$$

Da der Fehler F durch V und U vollständig bestimmt ist, gilt $H(F|V \otimes U) = 0$. Weiters gilt $H(F|V) \leq H(F) = H(p_e, 1 - p_e)$. Nun finden wir eine Abschätzung für $H(U|V \otimes F)$. Wegen der Kettenregel gilt

$$\begin{aligned} &H(U|V \otimes F) = H(U \otimes V|F) - H(V|F) \\ &= p_e \cdot H((U \otimes V)|_{F=1}) + (1-p_e) H((U \otimes V)|_{F=0}) - p_e H(V|_{F=1}) - (1-p_e) H(V|_{F=0}). \end{aligned}$$

Es gilt nun $H((U \otimes V)|_{F=0}) = H(V|_{F=0})$, da im Fall $F = 0$ die Ergebnisse von U und V übereinstimmen. Somit erhalten wir

$$H(U|V \otimes F) = p_e \cdot H(U|_{F=1} \,|\, V|_{F=1}) \leq p_e \cdot H(U|_{F=1}) \leq p_e \cdot \log_2(2^m) = p_e \cdot m.$$

Es gilt also

$$H(U|V) \leq H(p_e,\, 1-p_e) + p_e \cdot m$$

und somit

$$I(U;V) \geq H(U) - H(p_e,\, 1-p_e) - p_e \cdot m.$$

Wegen Satz 4.17 gilt $I(U;V) \leq I(\overline{X};\overline{Y})$. Wegen Satz 4.27 gilt $I(\overline{X};\overline{Y}) \leq n \cdot C$. Nach Voraussetzung gilt $H(U) = m$. Wir erhalten also

$$n \cdot C \geq m - H(p_e,\, 1-p_e) + p_e \cdot m,$$

also

$$\frac{n}{m} \cdot C \geq 1 - \frac{H(p_e,\, 1-p_e)}{m} - p_e,$$

und somit

$$p_e \geq 1 - \frac{C}{R} - \frac{H(p_e,\, 1-p_e)}{m}.$$

Wegen $H(p_e,\, 1-p_e) \leq 1$ folgt daraus direkt (1).

Für den Beweis von (2) definieren wir eine Zufallsvariable F durch

$$F := F_1 \otimes \cdots \otimes F_m.$$

Es gilt wieder

$$H(U \otimes F|V) = H(U|V) + H(F|V \otimes U)$$

und

$$H(U \otimes F|V) = H(F|V) + H(U|V \otimes F).$$

Da auch dieses F durch den Ausgang von U und V vollständig gegeben ist, gilt wieder $H(F|V \otimes U) = 0$. Ebenso ist U durch V und F vollständig bestimmt, also gilt $H(U|V \otimes F) = 0$. Sei p_i der Erwartungswert von F_i. Es gilt dann

$$H(F|V) \leq H(F) \leq \sum_{i=1}^{m} H(F_i) = \sum_{i=1}^{m} H(p_i,\, 1-p_i) = m \cdot \sum_{i=1}^{m} \frac{1}{m} H(p_i,\, 1-p_i).$$

Wegen der Konkavität der Funktion $x \mapsto H(x,\, 1-x)$ gilt

$$m \cdot \sum_{i=1}^{m} \frac{1}{m} H(p_i,\, 1-p_i) \leq m \cdot H(\frac{1}{m} \sum_{i=1}^{m} p_i,\, 1 - \frac{1}{m} \sum_{i=1}^{m} p_i).$$

Es gilt $\frac{1}{m} \sum_{i=1}^{m} p_i = \frac{1}{m} \sum_{i=1}^{m} E(F_i) = \frac{1}{m} E(\sum_{i=1}^{M} F_i) = p_B$. Insgesamt gilt also

$$H(F|V) \leq m \cdot H(p_B,\, 1-p_B).$$

Es gilt also

$$H(U|V) \leq m \cdot H(p_B,\, 1-p_B),$$

und somit wegen $H(U) = m$ auch

$$I(U;V) \geq m - m \cdot H(p_B,\, 1-p_B).$$

Da $I(U;V) \leq n \cdot C$, erhalten wir

$$n \cdot C \geq m - m \cdot H(p_B,\, 1-p_B),$$

und somit

$$\frac{C}{R} \geq 1 - H(p_B,\, 1 - p_B).$$

Wir beweisen nun (3). Wir definieren $F := \sum_{i=1}^{m} F_i$. Dann gilt Es gilt $p_B = \frac{1}{m}E(F) = \frac{1}{m}\sum_{j=1}^{m} j \cdot P[F = j] \leq \frac{1}{m}\sum_{j=1}^{m} m \cdot P[F = j] = \sum_{j=1}^{m} P[F = j] = P[F \geq 1] = p_e$ und $p_e = P[F \geq 1] = \sum_{j=1}^{m} P[F = j] \leq \sum_{j=1}^{m} j \cdot P[F = j] = E(F) = m\, p_B$. □

8. Sichere Übertragung

Ziel dieses Abschnittes ist, folgende Versions des Shannonschen Kanalkodierungssatzes zu beweisen.

Satz 4.31. *Sei* $\mathbf{K}$ *ein Kanal mit Kanalkapazität* C*, sei* $R < C$*, und sei* $\varepsilon > 0$*. Dann gibt es ein* $n_0 \in \mathbb{N}$*, sodass es für alle* $n \geq n_0$ *ein Übertragungssystem* $(U, \overline{X}, \overline{Y}, V)$ *mit* n*-facher Benutzung des Kanals* $\mathbf{K}$ *gibt, das folgende Eigenschaften erfüllt:*

(1) *Die Übertragungsrate* R' *erfüllt* $R' \geq R$*.*
(2) *Für alle Nachrichten* $\mathbf{x} \in \{0,1\}^{nR'}$ *gilt:* $P[V = \mathbf{x} \mid U = \mathbf{x}] \geq 1 - \varepsilon$*.*
(3) $\overline{Y}$ *ist eine zur* n*-fachen Verwendung des Kanals* $\mathbf{K}$ *und zu* $\overline{X}$ *passende Ausgabefolge.*

Wir werden den in [**CT06**] vorgestellten Weg einschlagen (siehe auch [**Mac03**]), und können diesen Satz nach einigen Vorbereitungen am Schluss dieser Sektion beweisen.

Seien $x_1, \ldots, x_M$ die Eingabezeichen und $y_1, \ldots, y_N$ die Ausgabezeichen dieses Kanals, für den wir ein Übertragungssystem suchen. Wir wählen s Elemente aus $\{x_1, \ldots, x_M\}^n$; mit diesen s Wörtern können wir s verschiedene Nachrichten, also ungefähr $\log_2(s)$ Bits kodieren. Für jedes dieser s Wörter $\mathbf{x}$ wählen wir eine Teilmenge B von $\{y_1, \ldots, y_N\}^n$, in der wir alle Wörter sammeln, die wir zu $\mathbf{x}$ decodieren.

Definition 4.32. Sei $((x_1, x_2, \ldots, x_M), (y_1, y_2, \ldots, y_N), A)$ ein Kanal, und seien $s, n \in \mathbb{N}$. Ein (s,n)*-Code für diesen Kanal* ist ein Paar

$$((\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(s)}), (B_1, \ldots, B_s)),$$

wobei für alle $i \in \{1, \ldots, s\}$ gilt: $\mathbf{x}^{(i)} \in \{x_1, x_2, \ldots, x_M\}^n$ und $B_i \subseteq \{y_1, y_2, \ldots, y_N\}^n$. Weiters fordern wir, dass für $i \neq j$ die Schnittmenge $B_i \cap B_j$ leer ist.

Wenn wir $\mathbf{y}$ empfangen und $\mathbf{y}$ liegt in B_j, dann decodieren wir $\mathbf{y}$ zu $\mathbf{x}^{(j)}$.

Wir überlegen uns nun, wie wahrscheinlich es ist, dass die Eingabe von $\mathbf{x}$ in den Kanal die Ausgabe $\mathbf{y}$ erzeugt.

DEFINITION 4.33. Sei $((x_1, x_2, \ldots, x_M), (y_1, y_2, \ldots, y_N), A)$ ein Kanal, und seien $\mathbf{x} \in \{x_1, x_2, \ldots, x_M\}^n$, $\mathbf{y} \in \{y_1, y_2, \ldots, y_N\}^n$. Dann definieren wir $w(\mathbf{y} \mid \mathbf{x})$ durch

$$w(\mathbf{y} \mid \mathbf{x}) = w((y_{j_1}, \ldots, y_{j_n}) \mid (x_{i_1}, \ldots, x_{i_n})) := \prod_{k=1}^{n} A(i_k, j_k).$$

Wir messen nun für jeden Code die Wahrscheinlichkeit für einen Decodierungsfehler. Da wir zunächst keine Wahrscheinlichkeitsräume definieren, verwenden wir das Wort *Plausibilität* und nicht das Wort *Wahrscheinlichkeit*.

DEFINITION 4.34. Sei $((x_1, x_2, \ldots, x_M), (y_1, y_2, \ldots, y_N), A)$ ein Kanal, seien $s, n \in \mathbb{N}$, und sei $((\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(s)}), (B_1, \ldots, B_s))$ ein Code für diesen Kanal. Dann definieren wir die *Fehlerplausibilität beim Senden von* $\mathbf{x}^{(i)}$ durch

$$e((\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(s)}), (B_1, \ldots, B_s), i) := 1 - \sum_{\mathbf{y} \in B_i} w(\mathbf{y} \mid \mathbf{x}^{(i)}).$$

Die *durchschnittliche Fehlerplausibilität* für diesen Code, abgekürzt $e((\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(s)}))$ ist gegeben durch

$$e((\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(s)}), (B_1, \ldots, B_s)) := \frac{1}{s} \sum_{i=1}^{s} e((\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(s)}), (B_1, \ldots, B_s), i).$$

DEFINITION 4.35. Seien $X : \Omega \to A$ und $Y : \Omega \to B$ Zufallsvariablen, sei $n \in \mathbb{N}$ und sei $\beta > 0$. Wir nennen ein $(\mathbf{a}, \mathbf{b})^T \in (A \times B)^n$ eine *gemeinsam typische Ausgangsfolge von* $X \otimes Y$ *der Länge* n *zum Parameter* β, wenn folgende Eigenschaften gelten:

(1) Die Folge $(\mathbf{a}, \mathbf{b})^T$ ist eine typische Ausgangsfolge von $X \otimes Y$ der Länge n zum Parameter β (wie in Defintion 3.39 definiert).
(2) Die Folge $\mathbf{a}$ ist eine typische Ausgangsfolge von X der Länge n zum Parameter β.
(3) Die Folge $\mathbf{b}$ ist eine typische Ausgangsfolge von Y der Länge n zum Parameter β.

Sei $GT(X, Y, n, \beta)$ die Menge der gemeinsam typischen Ausgangsfolgen von $X \otimes Y$.

PROPOSITION 4.36. *Seien* $X : \Omega \to A$ *und* $Y : \Omega \to B$ *Zufallsvariablen, sei* $n \in \mathbb{N}$*, und sei* $\beta > 0$*. Dann gilt*

$$P[(X \otimes Y)^{[n]} \in GT(X, Y, n, \beta)] \geq 1 - \frac{V(\mathrm{LogP}(X)) + V(\mathrm{LogP}(Y)) + V(\mathrm{LogP}(X \otimes Y))}{\beta^2 n}. \tag{4.5}$$

Beweis: Für die Wahrscheinlichkeit, dass $(\mathbf{a}, \mathbf{b})^T$ nicht typisch für $X \otimes Y$ zum Parameter β ist, gilt nach Proposition 3.40

$$P[(X \otimes Y)^{[n]} \notin T(X \otimes Y, n, \beta)] \leq \frac{V(\mathrm{LogP}(X \otimes Y))}{\beta^2 n}.$$

Für die Wahrscheinlichkeit, dass $\mathbf{a}$ nicht typisch für X zum Parameter β ist, gilt

$$P[X^{[n]} \notin T(X, n, \beta) \leq \frac{V(\mathrm{LogP}(X))}{\beta^2 n}.$$

Für die Wahrscheinlichkeit, dass $\mathbf{b}$ nicht typisch für Y zum Parameter β ist, gilt

$$P[Y^{[n]} \notin T(Y, n, \beta) \leq \frac{V(\mathrm{LogP}(Y))}{\beta^2 n}.$$

Die Wahrscheinlichkeit, dass zumindest eine der Eigenschaften dafür, dass $(\mathbf{a}, \mathbf{b})^T$ gemeinsam typisch für X und Y ist, nicht gilt, ist also höchstens

$$\frac{V(\mathrm{LogP}(X \otimes Y))}{\beta^2 n} + \frac{V(\mathrm{LogP}(X))}{\beta^2 n} + \frac{V(\mathrm{LogP}(Y))}{\beta^2 n}.$$

□

LEMMA 4.37. *Seien $X : \Omega \to A$ und $Y : \Omega \to B$ Zufallsvariablen. Wir definieren nun auf $\Omega^n \times \Omega^n$ die Zufallsvariablen $X^{[n]}_{allein}$ und $Y^{[n]}_{allein}$ durch*

$$X^{[n]}_{allein}(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2) := X^{[n]}(\boldsymbol{\omega}_1)$$

und

$$Y^{[n]}_{allein}(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2) := Y^{[n]}(\boldsymbol{\omega}_2).$$

Dann gilt $P[(X^{[n]}_{allein} \otimes Y^{[n]}_{allein})^T \in GT(X, Y, n, \beta)] \leq 2^{n(-I(X;Y)+3\beta)}$.

Beweis:

$$\begin{aligned} P[(X^{[n]}_{\mathrm{allein}} \otimes Y^{[n]}_{\mathrm{allein}})^T \in GT(X, Y, n, \beta)] &= \sum_{(\mathbf{a},\mathbf{b})^T \in GT(X,Y,n,\beta)} P[X^{[n]}_{\mathrm{allein}} = \mathbf{a}] \;\&\; P[Y^{[n]}_{\mathrm{allein}} = \mathbf{b}] \\ &= \sum_{(\mathbf{a},\mathbf{b})^T \in GT(X,Y,n,\beta)} P[X^{[n]} = \mathbf{a}] \cdot P[Y^{[n]} = \mathbf{b}]. \end{aligned}$$

Nun verwenden wir, dass $\mathbf{a}$ typisch für X und $\mathbf{b}$ typisch für Y ist und erhalten aus der Ungleichung (3.8)

$$\sum_{(\mathbf{a},\mathbf{b})^T \in GT(X,Y,n,\beta)} P[X^{[n]} = \mathbf{a}] \cdot P[Y^{[n]} = \mathbf{b}] \leq \sum_{(\mathbf{a},\mathbf{b})^T \in GT(X,Y,n,\beta)} 2^{-nH(X)+n\beta} \cdot 2^{-nH(Y)+n\beta}.$$

Da $GT(X, Y, n, \beta) \subseteq T(X \otimes Y, n, \beta)$, hat diese Summe wegen Proposition 3.41 höchstens $2^{nH(X\otimes Y)+n\beta}$ Summanden. Wir erhalten also

$$\sum_{(\mathbf{a},\mathbf{b})^T \in GT(X,Y,n,\beta)} 2^{-nH(X)+n\beta} \cdot 2^{-nH(Y)+n\beta} \leq 2^{n(H(X\otimes Y)-H(X)-H(Y)+3\beta)},$$

also insgesamt

$$P[(X^{[n]}_{\text{allein}} \otimes Y^{[n]}_{\text{allein}})^T \in GT(X,Y,n,\beta)] \leq 2^{n(-I(X;Y)+3\beta)}.$$

□

DEFINITION 4.38. Sei $\mathbf{K} = ((x_1, x_2, \ldots, x_M), (y_1, y_2, \ldots, y_N), A)$ ein Kanal, seien $s, n \in \mathbb{N}$, sei $\beta > 0$, und seien $X : \Omega \to \{x_1, x_2, \ldots, x_M\}$ und $Y : \Omega \to \{y_1, y_2, \ldots, y_N\}$ Zufallsvariablen, sodass Y eine zur Eingabe X und zum Kanal $\mathbf{K}$ passende Ausgabe ist. Für $\mathbf{a} \in \{x_1, x_2, \ldots, x_M\}^n$ definieren wir

$$J(\mathbf{a}) := \{\mathbf{y} \in \{y_1, y_2, \ldots, y_N\}^n \mid (\mathbf{a}, \mathbf{y}) \in GT(X,Y,n,\beta)\}.$$

Sei $\overline{\mathbf{x}} = (\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(s)}) \in (\{x_1, x_2, \ldots, x_M\}^n)^s$. Für $i \in \{1, \ldots, s\}$ definieren wir die Menge $B_i(X,Y,\overline{\mathbf{x}},\beta)$ durch

$$B_i\big(X,Y,\overline{\mathbf{x}},\beta\big) := J(\mathbf{x}^{(i)}) \setminus \bigcup \big\{J(\mathbf{x}^{(k)}) \mid k \in \{1,\ldots,s\},\, k \neq i\big\}.$$

Die Menge $B_i\big(X,Y,(\mathbf{x}^{(1)},\ldots,\mathbf{x}^{(s)}),\beta\big)$ enthält also alle $\mathbf{y}$, sodass $(\mathbf{x}^{(i)},\mathbf{y})$ gemeinsam typisch sind und $\mathbf{y}$ mit keinem anderen $\mathbf{x}^{(k)}$ gemeinsam typisch ist.

PROPOSITION 4.39. *Sei $\mathbf{K} = ((x_1, x_2, \ldots, x_M), (y_1, y_2, \ldots, y_N), A)$ ein Kanal, seien $s, n \in \mathbb{N}$, sei $\beta > 0$, und seien $X : \Omega \to \{x_1, x_2, \ldots, x_M\}$ und $Y : \Omega \to \{y_1, y_2, \ldots, y_N\}$ Zufallsvariablen, sodass Y eine zur Eingabe X und zum Kanal $\mathbf{K}$ passende Ausgabe ist.*

Die Zufallsvariable $X^{[n]}$ ist auf Ω^n definiert, die Zufallsvariable $(X^{[n]})^{[s]}$ auf $(\Omega^n)^s$. Sei F folgende auf $(\Omega^n)^s$ definierte Zufallsvariable:

$$F(\boldsymbol{\omega}) := e\big((X^{[n]})^{[s]}(\boldsymbol{\omega}), \big(B_1(X,Y,(X^{[n]})^{[s]}(\boldsymbol{\omega}),\beta),\ldots,B_s(X,Y,(X^{[n]})^{[s]}(\boldsymbol{\omega}),\beta)\big)\big).$$

Dann gilt

$$E(F) \leq \frac{V(\mathrm{LogP}(X \otimes Y)) + V(\mathrm{LogP}(X)) + V(\mathrm{LogP}(Y))}{\beta^2 \cdot n} + (s-1)\cdot 2^{n(-I(X;Y)+3\beta)}.$$

Beweis: Wir werden in den folgenden Rechnungen einige Abkürzungen verwenden.

$$\begin{aligned} \overline{\mathbf{x}} &:= (\mathbf{x}^{(1)},\ldots,\mathbf{x}^{(s)}) \\ p(\overline{\mathbf{x}}) &:= P[(X^{[n]})^{[s]} = \overline{\mathbf{x}}] \\ \mathcal{X} &:= \{x_1, x_2, \ldots, x_M\}^n \\ B_i(\overline{\mathbf{x}}) &:= B_i\big(X,Y,(\mathbf{x}^{(1)},\ldots,\mathbf{x}^{(s)}),\beta\big). \end{aligned}$$

Wir erhalten jetzt für den Erwartungswert von F:

$$\begin{aligned} E(F) &= \frac{1}{s}\sum_{i=1}^{s}\sum_{\overline{\mathbf{x}}\in\mathcal{X}^s} p(\overline{\mathbf{x}})\cdot e(\overline{\mathbf{x}},(B_1(\overline{\mathbf{x}}),\ldots,B_s(\overline{\mathbf{x}})),i) \\ &= \frac{1}{s}\sum_{i=1}^{s}\sum_{\overline{\mathbf{x}}\in\mathcal{X}^s} p(\overline{\mathbf{x}})\cdot\big(1-\sum_{\mathbf{y}\in B_i(\overline{\mathbf{x}})} w(\mathbf{y}\mid\mathbf{x}^{(i)})\big) \\ &= 1-\frac{1}{s}\sum_{i=1}^{s}\sum_{\overline{\mathbf{x}}\in\mathcal{X}^s} p(\overline{\mathbf{x}})\cdot\sum_{\mathbf{y}\in B_i(\overline{\mathbf{x}})} w(\mathbf{y}\mid\mathbf{x}^{(i)}). \end{aligned}$$

Sei S_i der i-te Summand der äußeren Summe. Dann gilt wegen der Definition der Mengen $B_i(\overline{\mathbf{x}})$

$$\begin{aligned} S_i &= \sum_{\overline{\mathbf{x}}\in\mathcal{X}^s} p(\overline{\mathbf{x}})\cdot\sum_{\mathbf{y}\in B_i} w(\mathbf{y}\mid\mathbf{x}^{(i)}) \\ &\geq \sum_{\overline{\mathbf{x}}\in\mathcal{X}^s} p(\overline{\mathbf{x}})\cdot\big(\sum_{\mathbf{y}\in J(\mathbf{x}^{(i)})} w(\mathbf{y}\mid\mathbf{x}^{(i)}) - \sum_{\substack{k=1\\k\neq i}}^{s}\sum_{\mathbf{y}\in J(\mathbf{x}^{(k)})} w(\mathbf{y}\mid\mathbf{x}^{(i)})\big) \\ &= \sum_{\overline{\mathbf{x}}\in\mathcal{X}^s} p(\overline{\mathbf{x}})\cdot\sum_{\mathbf{y}\in J(\mathbf{x}^{(i)})} w(\mathbf{y}\mid\mathbf{x}^{(i)}) \\ &\quad -\sum_{\overline{\mathbf{x}}\in\mathcal{X}^s}\sum_{\substack{k=1\\k\neq i}}^{s} p(\overline{\mathbf{x}})\cdot\sum_{\mathbf{y}\in J(\mathbf{x}^{(k)})} w(\mathbf{y}\mid\mathbf{x}^{(i)}). \end{aligned} \tag{4.6}$$

Wir berechnen als erstes den Ausdruck

$$\sum_{\overline{\mathbf{x}}\in\mathcal{X}^s} p(\overline{\mathbf{x}})\cdot\sum_{\mathbf{y}\in J(\mathbf{x}^{(i)})} w(\mathbf{y}\mid\mathbf{x}^{(i)}).$$

Wir fassen nun die Summanden mit gleichem $\mathbf{x}^{(i)}$ zusammen. Es gilt offensichtlich

$$\sum_{\overline{\mathbf{x}}\in\mathcal{X}^s} p(\overline{\mathbf{x}})\cdot\sum_{\mathbf{y}\in J(\mathbf{x}^{(i)})} w(\mathbf{y}\mid\mathbf{x}^{(i)}) = \sum_{\overline{\mathbf{x}}\in\mathcal{X}^s}\sum_{\mathbf{a}\in\mathcal{X}}\delta(\mathbf{a},\mathbf{x}^{(\mathbf{i})})\cdot p(\overline{\mathbf{x}})\cdot\sum_{\mathbf{y}\in J(\mathbf{x}^{(i)})} w(\mathbf{y}\mid\mathbf{x}^{(i)}).$$

Wir vertauschen die Summanden und erhalten

$$\sum_{\overline{\mathbf{x}}\in\mathcal{X}^s}\sum_{\mathbf{a}\in\mathcal{X}}\delta(\mathbf{a},\mathbf{x}^{(i)})\cdot p(\overline{\mathbf{x}})\cdot\sum_{\mathbf{y}\in J(\mathbf{x}^{(i)})} w(\mathbf{y}\mid\mathbf{x}^{(i)}) = \sum_{\mathbf{a}\in\mathcal{X}}\sum_{\overline{\mathbf{x}}\in\mathcal{X}^s}\delta(\mathbf{a},\mathbf{x}^{(i)})\cdot p(\overline{\mathbf{x}})\cdot\sum_{\mathbf{y}\in J(\mathbf{x}^{(i)})} w(\mathbf{y}\mid\mathbf{x}^{(i)}).$$

Für die Summanden, für die $\delta(\mathbf{x}^{(i)},\mathbf{a})\neq 0$, gilt $\mathbf{x}^{(i)}=\mathbf{a}$. Es gilt also

$$\sum_{\mathbf{a}\in\mathcal{X}}\sum_{\overline{\mathbf{x}}\in\mathcal{X}^s}\delta(\mathbf{a},\mathbf{x}^{(i)})\cdot p(\overline{\mathbf{x}})\cdot\sum_{\mathbf{y}\in J(\mathbf{x}^{(i)})} w(\mathbf{y}\mid\mathbf{x}^{(i)}). = \sum_{\mathbf{a}\in\mathcal{X}}\sum_{\overline{\mathbf{x}}\in\mathcal{X}^s}\delta(\mathbf{a},\mathbf{x}^{(i)})\cdot p(\overline{\mathbf{x}})\cdot\sum_{\mathbf{y}\in J(\mathbf{a})} w(\mathbf{y}\mid\mathbf{a}).$$

Da $\sum_{\mathbf{y}\in J(\mathbf{a})} w(\mathbf{y} \mid \mathbf{a})$ nicht von $\overline{\mathbf{x}}$ abhängt, können wir diesen Ausdruck herausheben und erhalten

$$\begin{aligned}
&\sum_{\mathbf{a}\in\mathcal{X}} \sum_{\overline{\mathbf{x}}\in\mathcal{X}^s} \delta(\mathbf{a}, \mathbf{x}^{(i)}) \cdot p(\overline{\mathbf{x}}) \cdot \sum_{\mathbf{y}\in J(\mathbf{a})} w(\mathbf{y} \mid \mathbf{a}) \\
&\qquad = \sum_{\mathbf{a}\in\mathcal{X}} \left(\sum_{\mathbf{y}\in J(\mathbf{a})} w(\mathbf{y} \mid \mathbf{a}) \right) \cdot \sum_{\overline{\mathbf{x}}\in\mathcal{X}^s} \delta(\mathbf{a}, \mathbf{x}^{(i)}) \cdot p(\overline{\mathbf{x}}) \\
&\qquad = \sum_{\mathbf{a}\in\mathcal{X}} \left(\sum_{\mathbf{y}\in J(\mathbf{a})} w(\mathbf{y} \mid \mathbf{a}) \right) \cdot P[X^{[n]} = \mathbf{a}] \\
&\qquad = \sum_{\mathbf{a}\in\mathcal{X}} \sum_{\mathbf{y}\in J(\mathbf{a})} P[X^{[n]} = \mathbf{a}] \cdot w(\mathbf{y} \mid \mathbf{a}).
\end{aligned}$$

Da die Zufallsvariable Y zum Kanal $\mathbf{K}$ und zur Eingabe X passend ist, gilt $P[X^{[n]} = \mathbf{a}] \cdot w(\mathbf{y} \mid \mathbf{a}) = P[X^{[n]} = \mathbf{a} \;\&\; Y^{[n]} = \mathbf{y}]$. Also gilt

$$\begin{aligned}
&\sum_{\mathbf{a}\in\mathcal{X}} \sum_{\mathbf{y}\in J(\mathbf{a})} P[X^{[n]} = \mathbf{a}] \cdot w(\mathbf{y} \mid \mathbf{a}) \\
&\qquad = \sum_{\mathbf{a}\in\mathcal{X}} \sum_{\mathbf{y}\in J(\mathbf{a})} P[X^{[n]} = \mathbf{a} \;\&\; Y^{[n]} = \mathbf{y}] \\
&\qquad = \sum_{(\mathbf{a},\mathbf{y})^T \in GT(X,Y,n,\beta)} P[X^{[n]} = \mathbf{a} \;\&\; Y^{[n]} = \mathbf{y}] \\
&\qquad = P[(X \otimes Y)^{[n]} \in GT(X, Y, n, \beta)].
\end{aligned}$$

Aus diesen Rechnungen und aus Satz 4.36 erhalten wir

$$\begin{aligned}
(4.7)\quad &\sum_{\overline{\mathbf{x}}\in\mathcal{X}^s} p(\overline{\mathbf{x}}) \cdot \sum_{\mathbf{y}\in J(\mathbf{x}^{(i)})} w(\mathbf{y} \mid \mathbf{x}^{(i)}) \\
&\qquad\qquad \geq 1 - \frac{V(\mathrm{LogP}(X)) + V(\mathrm{LogP}(Y)) + V(\mathrm{LogP}(X \otimes Y))}{\beta^2 n}.
\end{aligned}$$

Als nächstes nehmen wir $k \neq i$ an und schätzen den Ausdruck

$$\sum_{\overline{\mathbf{x}}\in\mathcal{X}^s} p(\overline{\mathbf{x}}) \cdot \sum_{\mathbf{y}\in J(\mathbf{x}^{(k)})} w(\mathbf{y} \mid \mathbf{x}^{(i)})$$

nach oben ab. Zuerst fassen wir wieder die Summanden mit gleichem $\mathbf{x}^{(k)}$ und $\mathbf{x}^{(i)}$ zusammen. Formal machen wir das so:

$$\begin{aligned}
&\sum_{\overline{\mathbf{x}}\in\mathcal{X}^s} p(\overline{\mathbf{x}}) \cdot \sum_{\mathbf{y}\in J(\mathbf{x}^{(k)})} w(\mathbf{y} \mid \mathbf{x}^{(i)}) \\
&\qquad = \sum_{\mathbf{a}\in\mathcal{X}} \sum_{\mathbf{b}\in\mathcal{X}} \sum_{\overline{\mathbf{x}}\in\mathcal{X}^s} \delta(\mathbf{a}, \mathbf{x}^{(i)}) \cdot \delta(\mathbf{b}, \mathbf{x}^{(k)}) \cdot p(\overline{\mathbf{x}}) \cdot \sum_{\mathbf{y}\in J(\mathbf{x}^{(k)})} w(\mathbf{y} \mid \mathbf{x}^{(i)}) \\
&\qquad = \sum_{\mathbf{a}\in\mathcal{X}} \sum_{\mathbf{b}\in\mathcal{X}} \sum_{\overline{\mathbf{x}}\in\mathcal{X}^s} \delta(\mathbf{a}, \mathbf{x}^{(i)}) \cdot \delta(\mathbf{b}, \mathbf{x}^{(k)}) \cdot p(\overline{\mathbf{x}}) \cdot \sum_{\mathbf{y}\in J(\mathbf{b})} w(\mathbf{y} \mid \mathbf{a}) \\
&\qquad = \sum_{\mathbf{a}\in\mathcal{X}} \sum_{\mathbf{b}\in\mathcal{X}} \left(\sum_{\mathbf{y}\in J(\mathbf{b})} w(\mathbf{y} \mid \mathbf{a}) \right) \cdot \sum_{\overline{\mathbf{x}}\in\mathcal{X}^s} \delta(\mathbf{a}, \mathbf{x}^{(i)}) \cdot \delta(\mathbf{b}, \mathbf{x}^{(k)}) \cdot p(\overline{\mathbf{x}}).
\end{aligned}$$

Da die s Komponenten von $(X^{[n]})^{[s]}$ unabhängig sind, ist dieser letzte Ausdruck gleich

$$\begin{aligned}
&\sum_{\mathbf{a}\in\mathcal{X}} \sum_{\mathbf{b}\in\mathcal{X}} \left(\sum_{\mathbf{y}\in J(\mathbf{b})} w(\mathbf{y} \mid \mathbf{a}) \right) \cdot P[X^{[n]} = \mathbf{a}] \cdot P[X^{[n]} = \mathbf{b}] \\
&\qquad = \sum_{\mathbf{b}\in\mathcal{X}} P[X^{[n]} = \mathbf{b}] \sum_{\mathbf{y}\in J(\mathbf{b})} \sum_{\mathbf{a}\in\mathcal{X}} P[X^{[n]} = \mathbf{a}] \cdot w(\mathbf{y} \mid \mathbf{a}) \\
&\qquad = \sum_{\mathbf{b}\in\mathcal{X}} P[X^{[n]} = \mathbf{b}] \cdot \sum_{\mathbf{y}\in J(\mathbf{b})} \sum_{\mathbf{a}\in\mathcal{X}} P[X^{[n]} = \mathbf{a} \;\&\; Y^{[n]} = \mathbf{y}] \\
&\qquad = \sum_{\mathbf{b}\in\mathcal{X}} P[X^{[n]} = \mathbf{b}] \cdot \sum_{\mathbf{y}\in J(\mathbf{b})} P[Y^{[n]} = \mathbf{y}] \\
&\qquad = \sum_{(\mathbf{b},\mathbf{y})^T \in GT(X,Y,n,\beta)} P[X^{[n]} = \mathbf{b}] \cdot P[Y^{[n]} = \mathbf{y}].
\end{aligned}$$

Wir definieren nun die Zufallsvariablen $X^{[n]}_{\text{allein}}$ und $Y^{[n]}_{\text{allein}}$ wie im Lemma 4.37. Dann erhalten wir

$$\begin{aligned}
&\sum_{(\mathbf{b},\mathbf{y})^T \in GT(X,Y,n,\beta)} P[X^{[n]} = \mathbf{b}] \cdot P[Y^{[n]} = \mathbf{y}] \\
&\qquad = \sum_{(\mathbf{b},\mathbf{y})^T \in GT(X,Y,n,\beta)} P[X^{[n]}_{\text{allein}} \otimes Y^{[n]}_{\text{allein}} = (\mathbf{b}, \mathbf{y})] \\
&\qquad = P[(X^{[n]}_{\text{allein}} \otimes Y^{[n]}_{\text{allein}})^T \in GT(X, Y, n, \beta)].
\end{aligned}$$

Aus diesen Rechnungen und Lemma 4.37 erhalten wir also

$$\sum_{\overline{\mathbf{x}}\in\mathcal{X}^s} p(\overline{\mathbf{x}}) \cdot \sum_{\mathbf{y}\in J(\mathbf{x}^{(k)})} w(\mathbf{y} \mid \mathbf{x}^{(i)}) \leq 2^{n(-I(X;Y)+3\beta)}. \tag{4.8}$$

Aus (4.6), (4.7) und (4.8) erhalten wir also

$$S_i \geq 1 - \frac{V(\mathrm{LogP}(X \otimes Y)) + V(\mathrm{LogP}(X)) + V(\mathrm{LogP}(Y))}{\beta^2 \cdot n} - (s-1) \cdot 2^{n(-I(X;Y)+3\beta)}.$$

Somit erhalten wir für $E(F)$:

$$\begin{aligned} E(F) &= 1 - \frac{1}{s} \sum_{i=1}^{s} S_i \\ &\leq \frac{V(\mathrm{LogP}(X \otimes Y)) + V(\mathrm{LogP}(X)) + V(\mathrm{LogP}(Y))}{\beta^2 \cdot n} + (s-1) \cdot 2^{n(-I(X;Y)+3\beta)}. \end{aligned}$$

□

Proposition 4.40. *Sei $((x_1, x_2, \ldots, x_M), (y_1, y_2, \ldots, y_N), A)$ ein Kanal mit Kanalkapazität C, und sei $\varepsilon > 0$. Sei $R < C$. Dann gibt es ein $n_0 \in \mathbb{N}$, sodass es für alle $n \geq n_0$ einen (s, n)-Code für diesen Kanal gibt, der folgende Eigenschaften erfüllt.*

1. $s = 2^{\lceil nR \rceil}$,
2. *Die durchschnittliche Fehlerplausibilität für diesen Code e erfüllt $e \leq \varepsilon$.*

Beweis: Wir setzen $\varepsilon_1 := \frac{C-R}{6}$, und $s := 2^{\lceil nR \rceil}$. Wir wählen als erstes Zufallsvariablen X, Y so, dass Y eine zur Eingabe X und zum Kanal $\mathbf{K}$ passende Ausgabe ist, und $I(X;Y) \geq C - \varepsilon_1$. Wir setzen $\beta := \varepsilon_1$. Die Zufallsvariable F aus Proposition 4.39 muss auch Werte annehmen, die nicht größer als ihr Erwartungswert sind. Es gibt also wegen Propsition 4.39 zumindest ein $\boldsymbol{\omega} \in (\Omega^n)^s$, sodass

$$F(\boldsymbol{\omega}) \leq \frac{V(\mathrm{LogP}(X \otimes Y)) + V(\mathrm{LogP}(X)) + V(\mathrm{LogP}(Y))}{\varepsilon_1^2 \cdot n} + (s-1) \cdot 2^{n(-I(X;Y)+3\varepsilon_1)}.$$

Wir wählen als Code

$$(\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(s)}) := (\mathbf{X}^{[n]})^{[s]}(\boldsymbol{\omega})$$

und

$$B_i := B_i(X, Y, (\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(s)}), \varepsilon_1).$$

Die Zahl $F(\boldsymbol{\omega})$ ist dann genau die durchschnittliche Fehlerplausibilität dieses Codes e. Es gilt nun

$$\begin{aligned} (s-1) \cdot 2^{n(-I(X;Y)+3\varepsilon_1)} &\leq 2^{\lceil nR \rceil} \cdot 2^{n(-I(X;Y)+3\varepsilon_1)} \\ &\leq 2^{nR+1} \cdot 2^{n(-C+\varepsilon_1+3\varepsilon_1)} \\ &= 2^{n(C-6\varepsilon_1+\frac{1}{n}-C+4\varepsilon_1)}. \end{aligned}$$

Wenn n so groß ist, dass $\frac{1}{n} \leq \varepsilon_1$, so gilt

$$2^{n(C-6\varepsilon_1+\frac{1}{n}-C+4\varepsilon_1)} \leq 2^{-\varepsilon_1 \cdot n}.$$

Wenn wir n nun so wählen, dass $n \geq \frac{1}{\varepsilon_1}$ und $\frac{V(\mathrm{LogP}(X\otimes Y))+V(\mathrm{LogP}(X))+V(\mathrm{LogP}(Y))}{\varepsilon_1^2 n} \leq \frac{\varepsilon}{2}$ und $2^{-\varepsilon_1 \cdot n} \leq \frac{\varepsilon}{2}$, so gilt $e = F(\boldsymbol{\omega}) \leq \varepsilon$. □

Wir können daraus jetzt das Übetragungssystem konstruieren.

Beweis von Satz 4.31: Wir erhalten aus Proposition 4.40 ein $n_0 \in \mathbb{N}$, sodass es für alle $n \geq n_0$ einen (s, n)-Code für den Kanal $\mathbf{K}$ mit $s = 2^{\lceil n\frac{C+R}{2}\rceil}$ gibt, dessen durchschnittliche Fehlerplausibilität höchstens $\frac{\varepsilon}{2}$ ist. Somit ist zumindest für die Hälfte der Codewörter die Fehlerplausibilität beim Senden des Wortes $\mathbf{x}^{(i)}$ höchstens ε. Wir nehmen an $(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(\frac{s}{2})})$ seien diese Wörter. Wir bilden daraus folgendes Übetragungssystem für Nachrichten der Länge $\lceil n\frac{C+R}{2}\rceil - 1$ mit n-maliger Verwendung des Kanals. Die Variable U wählt eine von $\frac{s}{2} = 2^{\lceil n\frac{C+R}{2}\rceil - 1}$ Nachrichten der Länge $\lceil n\frac{C+R}{2}\rceil - 1$, jede mit gleicher Wahrscheinlichkeit. Die Zufallsvariable $\overline{X}$ nimmt für die i-te Nachricht den Wert $\mathbf{x}^{(i)}$ an. Die Zufallsvariable $\overline{Y}$ ist zur Eingabe $\overline{X}$ und zur n-maligen Verwendung von $\mathbf{K}$ passend. Wir decodieren ein $\mathbf{y}$ zu i, falls $\mathbf{y} \in B_i$, und zu B_1, falls $\mathbf{y}$ in keinem B_i liegt. Die Zufallsvariable V nimmt dann als Wert die i-te Nachricht an. Die Fehlerwahrscheinlichkeit beim Senden der Nachricht i ist dann genau die Fehlerplausibilität beim Senden des $i$ten Codeworts (beziehungsweise sogar eventuell kleiner, falls $i = 1$). Es gilt (wir schreiben n_i für die i-te Nachricht):

$$\begin{aligned} P[V = n_i \mid U = n_i] &\geq P[\overline{Y} \in B_i \mid \overline{X} = \mathbf{x}^{(i)}] \\ &= \sum_{\mathbf{y}\in B_i} P[\overline{Y} = \mathbf{y} \mid \overline{X} = \mathbf{x}^{(i)}]. \end{aligned}$$

Da $\overline{Y}$ eine zur n-maligen Verwendung von $\mathbf{K}$ und $\overline{X}$ passende Ausgabefolge ist, gilt

$$\begin{aligned} \sum_{\mathbf{y}\in B_i} P[\overline{Y} = \mathbf{y} \mid \overline{X} = \mathbf{x}^{(i)}] &= \sum_{\mathbf{y}\in B_i} w(\mathbf{y} \mid \mathbf{x}^{(i)}) \\ &= 1 - e(\overline{\mathbf{x}}, \overline{B}, i) \\ &\geq 1 - \varepsilon. \end{aligned}$$

Nun berechnen wir die Übertragungsrate $R' = \frac{\lceil n\frac{C+R}{2}\rceil - 1}{n} \geq \frac{n\frac{C+R}{2} - 1}{n} = \frac{C+R}{2} - \frac{1}{n}$. Falls $n > \frac{2}{C-R}$, ist diese Übertragungsrate R' größer als R.

□

Literaturverzeichnis

[Ash90] R. B. Ash. *Information theory*. Dover Publications Inc., New York, 1990. Corrected reprint of the 1965 original.

[CT06] T. M. Cover and J. A. Thomas. *Elements of information theory*. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second edition, 2006.

[HQ95] W. Heise and P. Quattrocchi. *Informations- und Codierungstheorie*. Springer-Verlag, Berlin, third edition, 1995.

[Mac03] D. J. C. MacKay. *Information theory, inference and learning algorithms*. Cambridge University Press, New York, 2003. The book can be viewed at `http://www.inference.phy.cam.ac.uk/mackay/itprnn/book.html`.

[Wil99] W. Willems. *Codierungstheorie*. de Gruyter, Berlin, New York, 1999.